SINGLE CELL CHARACTERIZATION OF BIOCHEMICAL NOISE AND VARIABLE RESPONSE TO CHEMICAL INHIBITION

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Robert Michael Vogel

February 2016

SINGLE CELL CHARACTERIZATION OF BIOCHEMICAL NOISE AND VARIABLE RESPONSE TO CHEMICAL INHIBITION

Robert Michael Vogel, Ph.D.

Cornell University 2016

Individual cells utilize series of biochemical reactions, called signaling pathways, to translate environmental conditions to physiological responses. Consequently, the emergent properties of these signaling pathways are constrained to the physico-chemical laws of their biochemical constituents - they are strongly dependent on the number of molecular components per cell, intrinsically stochastic (noisy), and are inherently nonlinear. While these properties provide the plasticity required for a functioning living system, they present challenges for our understanding and control of cellular behavior. In this thesis I present single cell measurements (*i.e.* flow cytometry data) and physical models that we developed to track fluctuations in protein and phospho-protein abundance throughout biochemical reaction networks, and demonstrate how the nonlinear properties of biochemical reactions produce unique network responses to the targeted chemical inhibition of enzymes. We track the logarithmic fluctuations of biochemical components using a system of chemical Langevin equations and the corresponding Lyapunov equation. Used together, these equations uncover the connection between the organization of signaling pathway constituents and the covariance matrix estimated from the experimental data. With this formalism we theoretically explore the unique covariance representations of various signaling pathways, and experimentally validate our method in two established systems: a synthetic *E. coli* gene regulatory network and the Mitogen Activated Protein Kinase (MAPK) cascade in primary mouse T lymphocytes. In

addition, we use single cell measurements to mechanistically uncover the unique responses of signaling pathways, analog or digital, to targeted chemical inhibition. We extend these short time-scale properties of signaling pathways to a functional response, proliferation. Lastly, we show how the endogenous diversity of protein abundance among single cell clones provides a mechanism of resilience to chemical inhibition. Together, our combined experimental and theoretical approach provides novel insights to cellular systems, a method for directional inference, and optimal drug selection.

BIOGRAPHICAL SKETCH

Robert graduated from University of California Los Angeles in 2008 with a Bachelors of Science in Biophysics. He enrolled in the Physiology, Biophysics and Systems Biology program at Weill Cornell Medical College in 2010, eager to study under the supervision of Grégoire Altan-Bonnet.

ACKNOWLEDGEMENTS

I would like to thank my thesis adviser Grégoire Altan-Bonnet for his mentorship. Grégoire's mentorship consisted of frequent guidance, endless patience, and an inspiring dedication to research. I will always remember to pay the price of my models - that the predictions must be vetted by experiment.

Next, I would like to thank the members of the Altan-Bonnet Lab; both past and present. First Guillaume Voisinne for his guidance on model fitting, dynamic systems, how to ask thoughtful biological questions, and being a positive role model. I would specifically like to thank Karen Tkach and Jesse Cotari for mentoring me during my first years. I would then like to thank Alon Oyler-Yaniv, Jen Oyler-Yaniv, Ben Whitlock, and Amir Erez for their commentary and for constantly reminding me to remember the big biological questions.

Additionally, I would like to thank our collaborators at IBM: Bobby Prill, Guillermo Cecchi, and Gustavo Stolovitzky. I thoroughly enjoyed working with all of you and value your perspectives as theoreticians and guidance in studying stochastic processes in biological systems.

I also would like to thank former members of the Shimon Weiss lab at UCLA - specifically, Thomas Dertinger, Gopal Iyer, and Devdoot Majumdar. I will forever be indebted to you for your guidance during my undergraduate research experiences. From them, my interest in research grew, and my scientific standards defined.

Lastly, I would like to thank my family. First, Julia, for her limitless support, companionship, and tireless proof-reading. In addition, I would like to thank my parents and my sister for their support and 'interest' in what I was working on. Despite my excessive reliance on jargon they were always interested in my lab work. I would also like to thank my Aunt Nancy for inspiring me to study biology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Cancer represents a set of heterogeneous diseases that involve the pathological growth of cells within multicellular organisms. Despite the heterogeneity among types of cancers, these diseases share a common origin, that is the the accumulation of mutations and co-opted biological resources that manifest into six canonical behaviors, or 'hallmarks': sustained growth signals, evasion of growth suppressors, bypassing cell death mechanisms, unlimited reproductive capacity, reorganizing the micro-environment (*e.g.* angiogenesis), and metastasis [37]. As a result, disease management requires the selective disruption of these gained attributes in the pathological cells. While there have been many successful strategies introduced to treat cancer - general chemotherapies, targeted inhibitors (*e.g.* Imatinib [43, 44], Gefitnib [61], and Vemurafenib [18]), as well as immunological agents (*e.g.* monoclonal antibodies targeting either HER2 [9], Cytotoxic T-lymphocyte-associate Antigen 4 (CTLA4), and programmed cell death protein 1 (PD-1) [62]) - the complexity of selectively drugging tumors remains staggering.

One such complexity is the intrinsic heterogeneity in biological systems. Indeed, single cell studies have revealed that clonal cells respond variably to uniform perturbations [58, 21]. Often times exhibiting unique signal transduction responses to inhibitors [65, 75, 49, 12], diverse preferences for differentiation and proliferation [19, 79], and variable susceptibilities to death [7, 76]. Moreover, many of these phenotypes are not heritable to daughter cells. This lack of memory constrains the biological processes that can generate these observations. Mechanistic studies of cell behavior have produced promising progress by attributing the observed diversity to the physico-chemical properties of the cellular constituents [1, 80]. The

extent in which biological systems use these laws as a source of diversity and its application in disease remains to be determined.

1.1 Non-genetic phenotypic individuality, and maintenance of cellular diversity

Measurements that identified unique responses of individual cell clones to singular perturbations were pioneered in bacterial studies. An example is the elegant experiments of Balaban *et al* [7], who demonstrated that individual *E. coli* respond uniquely to treatment with Ampicillin. In their experiments they grew *E. coli* cells in small microfluidic channels that aligned the bacteria single file. They recorded the divisions of individual cells exposed to cycles of growth media and lethal doses Ampicillin. Upon Ampicillin administration a small population of slow growing cells remained and were designated 'persister cells'. Indeed, a possible mechanism of the apparent resilience of these persister cells is the presence of *de novo* mutations that conferred resistance to the antibiotic. However, this was not the case, upon reintroduction of Ampicillin the persister cell progeny were equally susceptible to the drug as the initial population, demonstrating that the persistence state was not heritable. These results suggest that the bacteria switched stochastically between a drug tolerant persistence state and a drug sensitive proliferative state. Indeed, stochastic switching of individual cells between states has been shown to be an effective mechanism of a population response to changing environments in systems where individuals do not have the capacity to measure environmental conditions [50, 1, 10].

In context of human disease, Gupta *et al* [36] showed similar behavior in human breast cancer derived cell lines. In theses cell lines, SUM149 and SUM159, the pop-

ulation of cells are composed of three cellular states, namely: basal, luminal, and stem-like. Presumably, each of these cell states represent stable populations, and that separation of these cells by fluorescence activated cell sorting (FACS) followed by independent cultures should create populations enriched for the respective phenotypes. However, after six days in culture, flow cytometry measurements revealed that each population reestablished the exact distribution of cell types as originally found in the parent culture. Indicating that diversity of cell states in cancer is not exclusively established by cancer stem cells. Moreover, similar findings were found in other mammalian cells. Mouse embryonic stem cells were found to stochastically maintain the frequency of Stella^{+}Pecam1^{+} and Stella^{-}Pecam1^{-}, markers correlative of hypothesized stem cell states, with a time scale of several days [38]. Similarly, mouse epithelial progenitor cells have been found to tune their respective proliferative versus differentiation bias in accordance with state of the system to maintain homeostasis [27]. While the utility in stochastic switching is clear, the actual molecular origins of the stochastic behavior are elusive.

Spudich and Koshland [77] were among the first to experimentally document and characterize cellular individuality. In their seminal work, they showed that individual bacteria exhibited unique flagellar response to chemoattractant. A seemingly unsurprising result considering the small abundance of each molecular component in the bacterial chemotaxis response pathway. Indeed, signaling pathways at this small scale are prone to stochasticity introduced from chemical reactions [34]. However, when they measured the individual responses of the bacteria they found that the individuality persisted for the lifetime of the bacterium. If indeed the individuality originated from the intrinsic fluctuations of chemical reactions, the measured behavior would persist for the lifetime of the chemical reactions, which is much shorter than the cell lifetime. A crucial fact that the authors used

to implicate protein variability as the origin of this non-genetic diversity [26].

A more direct linkage of phenotypic diversity to molecular components was found in a more complex setting, tumor necrosis factor related apoptosis-inducing ligand (TRAIL) induced apoptosis of HeLa cells [76]. These experiments consisted of tracking sister cells from birth to TRAIL induced mitochondria outer-membrane permeabilization (MOMP), a proxy for apoptosis and cell death, by time lapse fluorescent microscopy (Figure 1.1A). MOMP was measured by the release of IMS-RP, a fusion of red fluorescent protein (RFP) to a mitochondrial import sequence [2], from within the mitochondria to the cytosol. Cells were imaged and tracked for 20-30 hours prior to the treatment with exogenous TRAIL or TRAIL and cycloheximide, a potent inhibitor of protein synthesis. Figure 1.1B shows the correlation of individual sister cells time to MOMP with respect to their average life time. Immediately after cell division, the resulting two sister cells are effectively identical and as a result exhibit a high correlation in their times to TRAIL induced death. However, as the lifetime of the sister cells increases they become more dissimilar, which results in a small correlation in their death times. In cells treated with only TRAIL the correlation timescale, a measure of the memory, was only 1.5 hours. However, upon inhibiting the production of protein the correlation timescale increased to 11 hours (Figure 1.1B). Indeed over-expression of BID, a protein in the apoptotic regulatory signaling pathway, correlated with individual cell death times. These results suggest that protein production can function as an internal stochastic process that 'mixes' the time to death of individual cells.

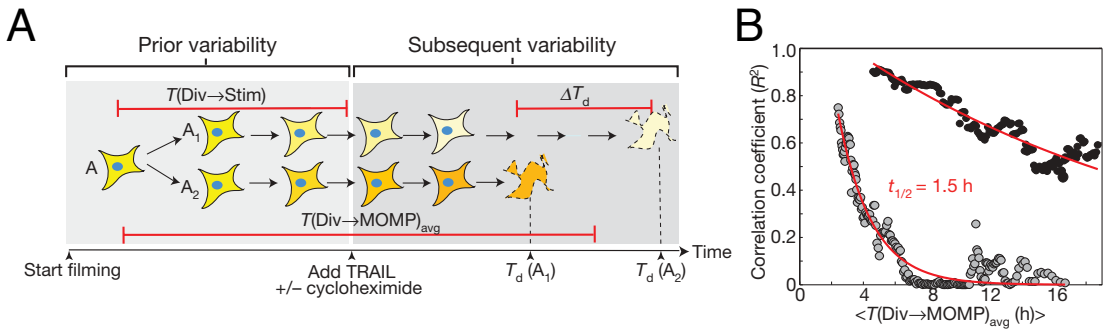

Figure 1.1: **Time to TRAIL induced apoptosis.** Reprinted by permission from Macmillan Publishers Ltd: Nature [76], copyright (2009). (A) Experimental design testing protein production variability as the mechanism for diverse death times of HELA cells treated with TRAIL. (B) Correlation of time to death among sister cells as a function of the average time between the mother's division and their respective time of MOMP.

1.2 Diversity of protein manifest to variable phenotypic responses

Unique phenotypic responses of individual cells has been correlated with the endogenous variability of protein in many biological settings. For example the variable abundance of the surface receptor Sca-1 in hematopoietic progenitor cells was shown to tune the differentiation rate into erythroid or myeloid lineages upon activation with Erythropoietin and GM-CSF, respectively [17]. Furthermore, they found that Sca-1 abundance per cell was stochastically maintained. Mechanisms to which the relative abundance of Sca-1 bias individual cell's differentiation rates are generally unknown. Although, an intuitive explanation is that Sca-1 diversity produces heterogeneous signaling response as implicated by Spencer *et al* in TRAIL induced apoptosis. Evidence of such a mechanism can be found in single cell measurements of AKT and ERK signaling in rat adrenal pheochromocytoma (PC12) cells. Classic experiments showed that administering epidermal growth factor (EGF) or neuronal growth factor (NGF) to PC12 cells induced proliferation

or differentiation into neurons, respectively [72]. This work was followed by single cell measurements of phospho-AKT (pAKT) and phospho-ERK (pERK) responses of PC12 cells to EGF or NGF stimulation. The pAKT and pERK plane identified a boundary of the phenotypic responses - pAKT signaling biased proliferation and pERK biased differentiation [19]. In addition, the response of cells to stimulation was often heterogeneous, the distribution of pAKT and pERK in single cells always spanned both sides of the boundary. As a result the respective stimulations modulated the fraction of cells exhibiting either phenotype. It remains to be shown if the abundance of Sca-1 in individual cells correlates with phospho-signaling responses that ultimately lead to unique phenotypic responses.

Independent measurements in T lymphocytes (T cells) has demonstrated the connection between protein expression variability and phospho-response diversity. One such example was found in peptide bound multi-Histocompatibility complex (pMHC) activation of the MAPK pathway in $CD8^+$ T cells. Single cell measurements simultaneously measuring both ppERK and CD8 demonstrated that the abundance CD8 correlated with the probability that a cell was activated. Moreover, measurements of the half effective quantity of antigen required for ERK activation (EC_{50}) revealed an anticorrelation with CD8. Complementing their measurements, detailed mechanistic models of TCR signaling reproduced the observed dependence of EC_{50} with CD8 [31]. Similarly, Cotari and Voisinne *et al* [22] demonstrated 100 fold change in the sensitivity of STAT5 phosphorylation among clones responding to Interleukin 2 (IL-2) according to the relative abundance of the IL2 receptor α chain (IL2Rα) per cell. They followed this observation by showing how cytokine receptors compete for a common component, namely the common γ chain. Consequently, the abundance of IL2Rα tuned the responsiveness of cells signaling through the IL-7 pathway by sequestering the common γ chain from the

IL7 receptor α chain, and as a result inhibits IL-7 signaling. Interestingly, a separate study showed that reintroduction of $CD8^+$ T cells sorted by their IL2Rα expression (low, hi) into an LCMV infected mouse developed different frequencies of memory and effector cells [41]. It is tempting to attribute this result to the work of Cotari and Voisinne *et al.* Indeed, protein variability is an attractive mechanism to break symmetric responses of individual clones to stimuli.

The composite of all these experimental observations - that protein abundance is diverse among single cell clones, tunes the responsiveness of signal transduction pathways and lastly bias cell fate - suggest that the endogenous variability of protein is an important property of biological systems. However, it is not clear what mechanisms can biological systems utilize to generate the observed diversity. Specifically, does the diversity arise from very subtle differences in the cell's environment (extrinsic noise in gene expression)? Or, does the variability manifest from internal processes (intrinsic noise in gene expression), would cells in a 'perfectly' homogeneous environment exhibit unique phenotypes? Understanding the mechanisms of protein abundance variability would identify the extent in which biological systems can rely on these tools to establish diversity.

1.3 The physical origins of gene expression variability

The theoretical tools to study the intrinsic properties of protein abundance fluctuations preceded experimental capabilities by more the twenty five years [84, 34]. As a result, there are numerous results that characterize the magnitude of protein abundance fluctuations [26, 13], numerical analysis of the propagation of protein fluctuations through gene regulatory networks [54], and establishment of funda-

mental limits to which negative regulation can decrease undesired variability [63]. Of particular interest is Thattai and van Oudenaarden's [82] development of a simple and analytical framework to study the magnitude of protein fluctuations for various gene regulatory networks.

In this work Thattai and van Oudenaarden *assumed* that the fluctuations in protein abundance are the manifestation of the intrinsically stochastic properties of the chemical processes comprising the central dogma of molecular biology [23]. Indeed, the canonical mathematical representation of these processes is the Master equation, which attributes the time derivative of the probability in observing n molecules to the sum of the production and degradation of the probabilistic processes. In this case, there are three biological species they keep track of: DNA, RNA, and protein, which results in the Master equation

$$\frac{d}{dt}f(\vec{n},t) = \sum_{j=1}^{N}\left\{E_j^{+1}\lambda_j n_j f(\vec{n},t) + \left(E_j^{-1}-1\right)\sum_{i=1}^{N} a_{ji}n_i f(\vec{n},t)\right\} - \sum_{i=1}^{N}\lambda_i n_i f(\vec{n},t), \tag{1.1}$$

where $f(\vec{n},t)$ represents the joint probability of observing n_i copies of the i^{th} biochemical species for all $i = \{1,2,\ldots,N\}$ at time t. The operators $E_j^{\pm}$ represents van Kampen's [84] translational operator, which takes $E_j^{\pm}n_j f(\vec{n}) \rightarrow (n_j \pm 1)f(n_1,\ldots,n_j \pm 1,\ldots n_N)$. Lastly, the constants a_{ji}'s represent the response of species j to species i, $a_{jj} = 0$ for all j, and λ_i's are decay constants. Solving Equation 1.1 for $f(\vec{n},t)$ is often a difficult task, and is made easier by exploiting the properties of generating functions.

Generating functions are standard approach for solving Master equations when the coefficients are linear with respect to the abundance of each variable n_i (remember n_i counts molecules, therefore they are integers). The goal of the generating

function is to recast the problem into a partial differential equation with respect to new continuous variables z_i for each biochemical species and the generating function F. By switching to this representation one can now utilize the many toolboxes for solving PDE's to find the desired probability distribution. However, there is an alternative benefit of using generating functions. That is, they provide a method for assessing the moments of the the corresponding stochastic process. This problem is often far easier than solving for the entire probability distribution. Thattai and van Oudenaarden [82] used this approach to compute the fluctuations. First, they computed the expression for the generating function that corresponded to their stochastic representation of the biology's central dogma,

$$\frac{\partial F}{\partial t} = \sum_{j=1}^{N}(1 - z_j)\left(\lambda \frac{\partial F}{\partial z_j} - \sum_{i=1}^{N} a_{ji} z_i \frac{\partial F}{\partial z_i}\right). \tag{1.2}$$

To continue, they set the partial derivative with respect to time to zero - they only cared about the steady state moments. Next, Taylor expanding this expression with respect to all z_i's about z_i's = 1, provided the moment generating function of the steady state probability distribution. From this they developed a general expression for the steady state averages ($\mathbf{J}$) and the covariances ($\mathbf{K}$),

$$(\mathbf{A} - \mathbf{\Lambda})\mathbf{J} = 0, \tag{1.3a}$$

$$(\mathbf{A} - \mathbf{\Lambda})\mathbf{K} + \mathbf{L} = -[(\mathbf{A} - \mathbf{\Lambda})\mathbf{K} + \mathbf{L}]^T, \tag{1.3b}$$

where $\mathbf{\Lambda}$ is a diagonal N x N matrix of λ_i's, $\mathbf{A}$ is the matrix connectivity that describes the interacting biochemical species, $\mathbf{J}$ is the N x 1 vector of averages, $\mathbf{K}$ is the N x N covariance matrix, and lastly the elements of $\mathbf{L}$ are $L_{ji} = a_{ji} J_i$. Therefore, by constructing the matrix $\mathbf{A}$ and substituting it into Equation 1.3 they are able to compute the statistics of each biochemical species of the model.

Figure 1.2 represents their model characterizing the statistics of the central

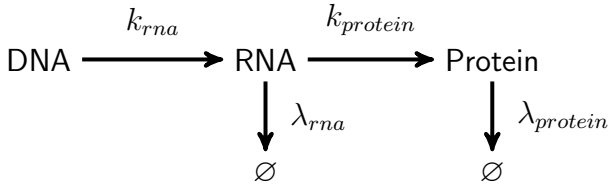

Figure 1.2: **Modeling the central dogma of molecular biology.** The model and corresponding parameters analyzed by Thattai and van Oudenaarden [82]

dogma. The mathematical representation of this process takes the form of,

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ k_{rna} & 0 & 0 \\ 0 & k_{protein} & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Lambda} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda_{rna} & 0 \\ 0 & 0 & \lambda_{protein} \end{bmatrix}, \tag{1.4}$$

which can then be substituted into Equation 1.3 to find the mean and variance of this process. In this model, the top row represents the dynamics of DNA abundance, which doesn't change (*i.e.* no cell cycle), the second row represents the dynamics of RNA, and lastly the third row represents the dynamics of protein production. The results are reported as the Fano factor, a quantity that characterizes the variance of the distribution normalized to the mean. This quantity is ideal for contextualizing properties of the fluctuations in protein abundance - *i.e.* identifies sets of kinetic parameters that result in exceedingly variable protein abundance given the mean expression. As a result, this quantity is important to consider in biological systems that tune their gene expression parameters to maximally or minimally diversify their population. Furthermore, if protein abundance fluctuations are small and behave according to Poisson statistics the Fano factor will be unity. The Fano factor for the fluctuations of protein abundance according to this model is,

$$\frac{\sigma^2_{protein}}{\mu_{protein}} = \frac{b}{1+\gamma} + 1 \approx b + 1, \tag{1.5}$$

where $b = \frac{k_{protein}}{\lambda_{rna}}$, and $\gamma = \frac{\lambda_{protein}}{\lambda_{rna}}$. Burst size, b, represents the number of proteins translated per RNA transcript, and γ quantifies the separation of RNA and

protein dynamic timescales. Equation 1.5 represents the success of the analysis, by specifically attributing the fluctuations in protein abundance in physiological conditions ($\lambda_{protein} \ll \lambda_{rna}$) to the burst size. Moreover, the fluctuations are larger than those estimated by Poisson statistics, and by tuning the kinetic parameters of b individual cells can tune the diversity of protein abundance among clones. Furthermore, Equation 1.5 provides the functional dependence of the fluctuations to the separation of RNA and protein degradation timescales. Indeed, if physiological conditions were such that $\lambda_{protein} \gg \lambda_{rna}$ then, the protein levels would exactly reflect the RNA abundance. Considering this scenario in Equation 1.5 by setting $\gamma \to \infty$ we see that the fluctuations converge to the Poisson limit. Taken together, cells can carefully tune both γ and b to create more or less diverse populations.

Cells can also tune their gene products by constructing feedback regulated circuits. Indeed, negative feedback regulation is an important mechanism to tune the probability distribution of protein products [63]. To incorporate this mechanism into their analytic expression (Eq 1.3), Thattai and van Oudenaarden took

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ k_{rna} & 0 & -k_{rna}^{feedback} \\ 0 & k_{protein} & 0 \end{bmatrix}, \tag{1.6}$$

while not changing $\mathbf{\Lambda}$. The corresponding Fano factor is,

$$\frac{\sigma^2_{protein}}{\mu_{protein}} = \left(\frac{1-\phi}{1+b\phi}\right)\left(\frac{b}{1+\gamma}\right) + 1, \tag{1.7}$$

where $\phi = \frac{k_{rna}^{feedback}}{\gamma_p}$ and represents the feedback strength. Indeed, comparing the Fano factors of the feedback regulated motif (Eq 1.7) to that of the unregulated protein synthesis model (Eq 1.5) reveals the advantages of feedback. By tuning the strength of the feedback, ϕ, cells can precisely tune the fluctuations in protein abundance to levels below the Poisson limit, a result corroborated by many other studies [63, 81, 73].

Extending these results to the observed diversity in biological systems is particularly intriguing. The ability of cells to tune their protein abundance fluctuations suggests that in some cases they intentionally diversify by tuning the kinetic parameters of protein synthesis. In contrast, important regulatory pathways such as apoptosis need to be tightly controlled, which may explain why over-expression studies in HeLa cells were required to correlate variability in cells time to death [76]. Yet, despite the elegance of their solution, and its obvious implications to the numerous accounts of biological diversity, the relevance of intrinsic fluctuations had yet to be vetted.

Advances in synthetic biology and fluorescent microscopy provided the tools required to separate the mechanisms of protein variation among single cells (intrinsic and extrinsic noise). Elowitz *et al* [28] decomposed these sources of variability by making strains of *E. coli* that encode both *cfp* and *yfp* under identical promoters and integrated symmetrically about the origin of replication. By measuring each fluorescent protein in individual cells they were able to estimate the contributions of both sources of diversity. They reported that indeed individual cells exhibited variation in protein abundance that both correlated (extrinsic) and did not correlate (intrinsic) with one another. Moreover, they reported that the square of the coefficient of variation depreciated with increased promoter activity. In context of Thattai and van Oudenaarden's analysis increased promoter activity is represented by an increase in k_{rna}. Computing the corresponding coefficient of variation,

$$\mathrm{CV}^2 = \frac{\lambda_{protein}}{k_{rna} b} \left[\frac{b}{1+\gamma} + 1 \right], \tag{1.8}$$

which shows the explicit dependence of the promoter strength in reducing the measured coefficient of variation. This result published by Elowitz *et al.* [28] was the first experimental confirmation that fluctuations of protein in biological systems can originate from the statistics of cell internal processes.

Indeed, Elowitz *et al.* experimentally vetted the fundamental assumption of Master equation description of protein synthesis. The next important attribute to consider is the prediction that protein production burst exists and dominate the variability. Measurements of this process were beautifully conducted with single *E. coli* cells in microfluidic chambers (volume $\approx$ 100 pL, Fig 1.3A) and media supplemented with fluorescein-di-β-D-galactopyranoside (FDG, Fig 1.3C) [16]. FDG is enzymatically hydrolyzed by β-galactosidase (β-gal), an enzyme produced by *E. coli* to digest lactose and other complex sugars, into galactose and the fluorescent molecule fluorescein. Fluorescein in *E. coli* is quickly pumped out of the cell, and without the small volume of the microfluidic device would be unmeasurable by fluorescence microscopy. Real time fluorescent measurements of fluorescein abundance revealed that the time derivative of fluorescein accumulation increased in jumps (Figure 1.3B,D). The magnitude of these jumps are proportional to the number of β-gal molecules produced per burst of protein synthesis - the net speed of FDG hydrolysis is proportional to the number of enzymes present. A histogram of the number of β-gal enzymes produced per burst uncovered an exponential distribution with on average 5 proteins produced per burst (Figure 1.3E). The shape of the distribution is reflective of the exponential life-time of mRNA in the cells, and is in complete agreement with the model assumptions of Thattai and van Oudenaarden [82].

With these experimental facts established, Friedman *et al.* [33] extended the work of Thattai and van Oudenaarden by deriving an analytical distribution of protein abundance per cell. They found that protein abundance per cell could be approximated by a continuous and long-tailed Gamma distribution. Indeed, subsequent analysis confirmed that long-tailed distribution of protein is universal across all phylogenies [68, 70]. Together, these results show that protein variability

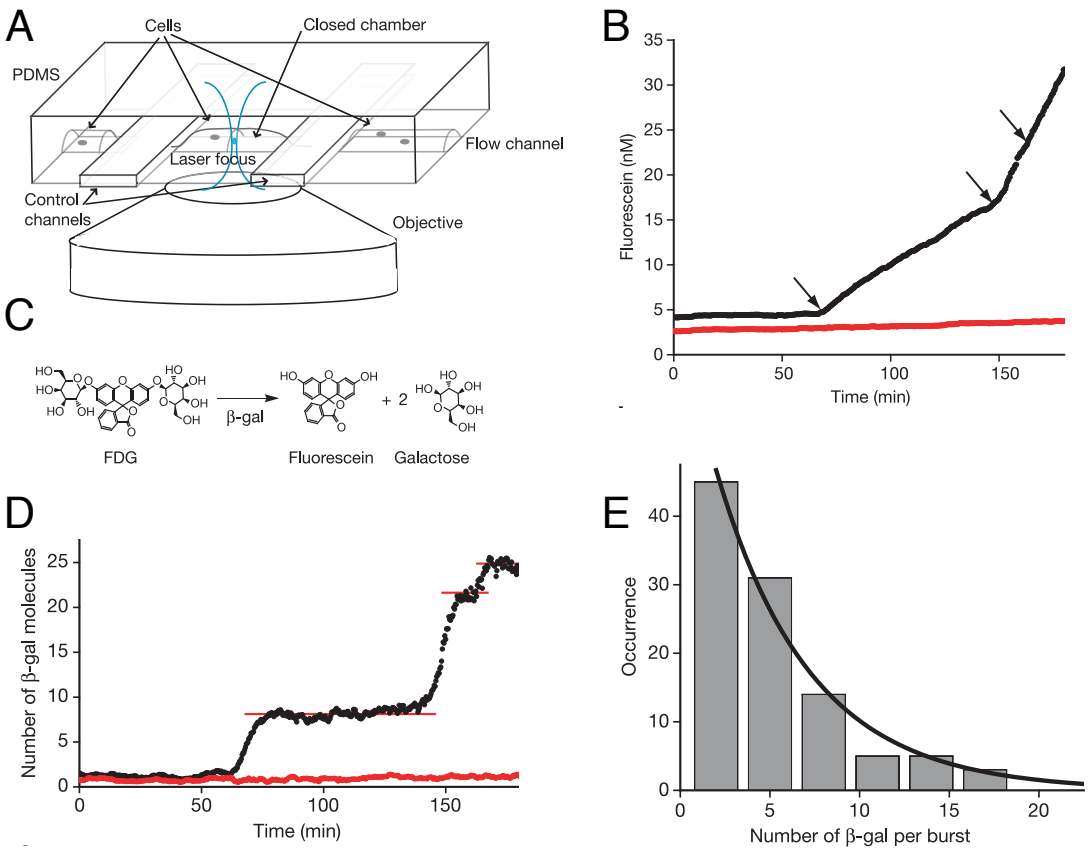

Figure 1.3: **Protein production burst statistics.** Reprinted by permission from Macmillan Publishers Ltd: Nature [16], copyright (2006). (A) The microfluidic experimental apparatus. (B) Real time measurements of fluorescein per micro-well. (C) β-galactosidase hydrolysis of FDG into galactose and fluorescein. (D) Real time measurements of abundance of β-gal per micro-well. (E) Histogram of β-gal protein produced per burst and the corresponding exponential fit.

is a stochastic processes, and that the magnitude of the resulting protein diversity is tunable by both the kinetic parameters of gene expression and the structure of gene regulatory networks.

1.4 Stochastic gene expression correlates with phenotypic variability

The intrinsic stochasticity of gene expression is an elegant mechanism proposed to generate the behavioral individuality observed in both prokaryotes and eukaryotes. Indeed, numerical simulations of gene regulatory network activity suggest that the stochastic expression of protein is sufficient to generate variable biological behaviors. For example, simulations of gene expression in *E. coli* infected with phage λ accurately predicted the fraction of cells exhibiting either lysogenic or lytic phenotypes [6]. Moreover, diversity originating from the stochastic accumulation of IL2Rα was sufficient in generating diverse times of cell cycle entry in T lymphocytes [87]. Additionally, the stochasticity in gene expression in mouse Embryonic Stem cells generated variability in the dynamics of the Oct4, Sox2, and Nanog gene regulatory network. This variability was shown to give rise to a transient Nanog low state capable of differentiation [42]. Indeed, these results suggest that gene expression variability is sufficient to generate unique biological behaviors, yet they do not provide unequivocal evidence for this mechanism. In this section we present an examples in which biological individuality is generated by the diversity of protein products.

Korobkova *et al.* [48] measured bacterial chemotactic response with real time single cell measurements of flagellar activity in engineered strains of *E. coli*. Here, they monitored the rotation preferences, clockwise (CW) or counter-clockwise (CCW), of individual and immobilized *E. coli* as a measure of the chemotactic signaling pathway response. The temporal data were then encoded in binary (CW or CCW) and were characterized by the the corresponding power spectrum. The power spectrum is a sensitive measure of the magnitude of correlations at different

frequencies, inverse time lags. Strikingly, they observed that the power spectrum of unstimulated *E. coli* flagellar motion peaked at 10^{-3} Hz, a large deviation from the timescale representative of the Poisson process (0.1Hz - 1 Hz) of this network of chemical reactions. Indeed, these fluctuations arise from the complex nonlinear properties of the molecular reactions in this network. Interestingly, they found that by changing the internal concentration of the regulator CheR, they were able to tune the correlation timescale of the bacteria. If we consider a population of clones of bacteria, the stochastic production of CheR would tune their respective CW biases. Unique CW biases would macroscopically manifest as unique individual response times to stimulus as first observed by Spudich and Koshland [77].

Another important behavior of cells is the transient switching of differentiation states. Süel *et al.* [79], demonstrated that the bacteria *B subtilis* utilizes the fluctuations of protein to transiently switch from a vegetative state to a transient competent state. In this system the switch to a competent state occurs upon upregulation of the master regulator ComK and its corresponding gene products. *comK* is constitutively expressed at low levels during the stationary phase of the bacteria, but can't initiate competency due to MecA catalyzed degradation of the protein. ComS, another protein involved in this gene regulatory network functions as an inhibitor of MecA (Figure 1.4A), and as a result ComK increases and initiates competency. To understand how these complex dynamics manifest into a transient differentiation state, Süel *et al* concomitantly monitored the fluorescence of CFP and YFP by quantitative fluorescence microscopy in individual bacteria. *cfp* expression was upregulated by the presence of ComK, meanwhile *yfp* expression was controlled by the abundance of ComS. Figure 1.4B shows the resulting dynamics of these two fluorescent proteins in a single bacteria. Here, we see that the dynamics of ComK follow those of ComS and as a result exhibit staggered limit cycles.

Combining these observations and the interactions of the gene regulatory network structure they built a nonlinear dynamical model of ComS and ComK abundance. The corresponding dynamics of this model are summarized by the phase plane in Figure 1.4C. In this figure we see that the nullclines representing the stationary dependency of ComK (blue) and ComS (green) cross in three places, and represent three stationary or fixed points in the dynamical system. The system rests in the vegetative state which is represented by the stable fixed point (filled circle). However, upon experiencing a large fluctuation in the expression of *comS*, the system must cycle about the remaining two unstable fixed points (open circles) to return to the stable fixed point. This transient and stochastically induced cycle represents the transition to the transient competent state. Indeed, the stochastic expression of protein was required to initiate the dynamics of competency.

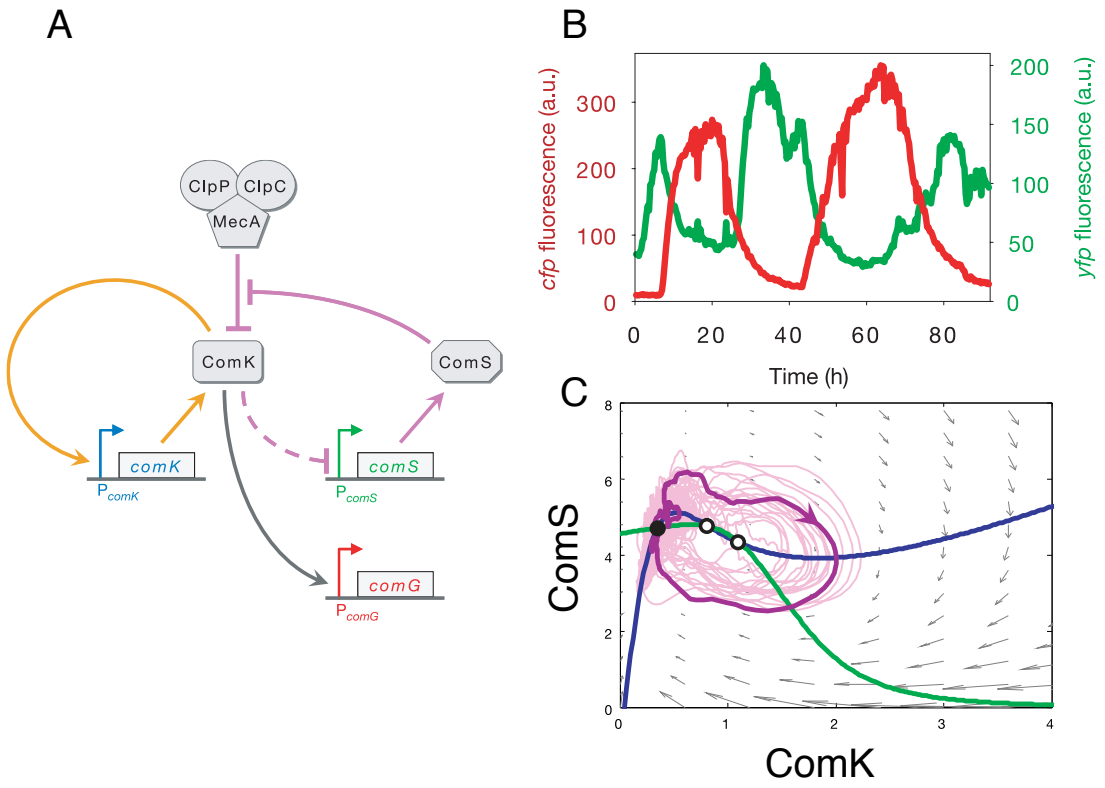

Figure 1.4: **Excitable gene regulatory network controls transient differentiation.** Reprinted by permission from Macmillan Publishers Ltd: Nature [79] copyright (2006). (A) The simplified gene regulatory network controlling the transient competent state in *B subtilis*. (B) The dynamics of ComK (CFP) and ComS (YFP) abundance in a single cell. *cfp* and *yfp* expression was encoded under promoters activated by ComK and ComS, respectively. (C) The phase diagram summarizing the dynamics of ComK and ComS. Blue and green lines represent the nullclines from the dynamic equations of ComK and ComS, respectively. The purple line represents a single representative competency cycle. Pink lines are samples of stochastic trajectories in the ComK, ComS plane. The arrows show the relative strengths of the velocity field.

1.5 In this dissertation

Single cell measurements have revealed that biological systems are diverse and complex. In many cases this diversity is driven by the stochastic behavior of chemical reactions in the limit of small numbers of molecules. Exacerbating these statistical fluctuations in molecular abundance of biochemicals are the complex nonlinear processes they compose, often capable of explosive amplification from small differences in initial conditions. Because, these phenomena are general properties of the tools at the disposal of cells it is imperative to theoretically understand the space of possible solutions these systems can explore and validate predictions with experiments at the appropriate resolution, the single cell. In this dissertation I present work, done in collaboration with colleagues at Memorial Sloan Kettering Cancer Center and at IBM, that aims in providing insight to properties of stochastic biological process (*e.g.* gene expression and phospho-signaling) and the consequences of diversity and nonlinear structure in chemical inhibition.

CHAPTER 2

PROPAGATION OF FLUCTUATIONS IN THE ABUNDANCE OF BIOMOLECULES

Cells make decisions in response to environmental stimuli by using systems of biochemical reactions. These reactions are required to quickly infer the cell's environmental state, and translate that information into an executable cellular program. Successful execution of a program consists of activating the optimal set of genes required for the desired response. The accurate response to the environment is predicated on accurate processing by the network of biochemical reactions. However, these reactions are subject to randomness introduced by the discrete nature of molecules. Indeed, simulations of chemical reactions with small numbers of reactants manifest in products that fluctuate in their concentration [84, 34, 35]. In these simulations one assumes that the fluctuations are exclusively attributed to the infrequent random collisions of molecules of well mixed idealized systems. Cellular systems are often the opposite; they are crowded, not well mixed, and do not always consist of a small number of molecules. Consequently, several pioneering studies have shown how biological systems exhibit noisy processing beyond that predicted by Poisson processes [82, 33, 16]. However, there is no systematic analysis of the transmission of fluctuations in the abundance of biomolecules originating from such observed diversity.

In this chapter I present a study of biochemical noise conducted in collaboration with members of IBM T.J. Watson Research Center [66]. Our goal was to develop a framework to infer the structure of the biochemical network that generates a measured response to perturbation by analyzing the fluctuations (so-called noise) in protein abundance at steady state. In this study we characterized the fluctuations

in the abundance of biomolecules in two biological systems: 1) gene expression in a synthetic gene regulatory network in *E. coli* measured and previously published by Pedraza and van Oudenaarden [64], and 2) in the phospho-signaling of primary T lymphocytes. We used single cell measurements of protein and phospho-signal abundance to identify the steady state distribution which the fluctuations are sampled from. Using our experimentally motivated distribution we developed a chemical Langevin equation and investigated the properties of the fluctuations in biochemical abundance with respect to three network configurations. Lastly, we used our framework to infer the underlying network configuration associated with signal propagation in our experimental system. We used the consistency of our experimental findings with current understanding of the biological systems under study (by design, for the synthetic circuit; by common understanding, for the signal transduction cascade) to validate our approach.

2.1 Fluctuations in biochemical abundance is log-normal

In this section, we introduce the two biological systems we investigated, at the single cell level, in order to tease out the role and significance of biochemical fluctuations.

2.1.1 Fluctuations in a synthetic *E. coli* gene regulatory network

Pedraza and van Oudenaarden [64] developed a simple synthetic gene regulatory network to measure the propagation of protein fluctuations from one gene to another in *Escherichia coli*. This network utilized four genes designated Gene 0 (G0),

Gene 1 (G1), Gene 2 (G2) and lastly Gene 3 (G3). In this small network G0 inhibits the expression of G1, G1 inhibits the expression of G2, and G3 does not interact with either of the genes (Figure 2.1). As a result, any correlation of the abundance of Genes 0,1,2 with Gene 3 originate from extrinsic sources of variability (*e.g.* variability in plasmid copies per cell). To eliminate this source of variability, we normalized the fluorescence representing the abundance of Gene 1 and Gene 2 in each cell by that of Gene 3.

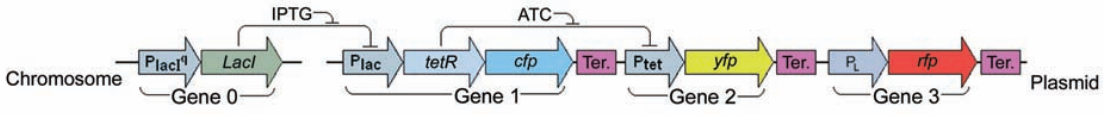

Figure 2.1: **Synthetic *E. coli* gene network.** From [64]. Reprinted with permission from AAAS. The synthetic gene regulatory network developed by Pedraza and van Oudenaarden.

The first three genes in the network form a cascade of negative regulators (Figure 2.1). The first gene in the cascade, G0, constitutively expresses the *lacI* gene which encodes the lactose repressor, an inhibitor of the lac promoter. In the absence of lactose, or the lactose analog isopropyl-β-D-thiogalactopyranoside (IPTG), the lac repressor inhibits the transcription of G1. As a result, the exogenous administration of IPTG provides an experimental tool for tuning the activity of the network. G1 encodes for both *tetR* and *cfp* genes, which encode for the tetracycline repressor and the cyan fluorescent protein (CFP), respectively. Continuing down the cascade, the tetracycline repressor inhibits the activation of the tetracycline promoter, which in turn controls the transcription of G2 and consequently the abundance of the yellow fluorescent protein (YFP). The inhibitory capacity of the tetracycline repressor can be modulated by the administration of exogenous anhydrotetracycline (ATC), however we do not investigate the consequences of ATC administration.

The fluctuations of protein abundance for each gene product pair (CFP, YFP) were monitored in individual *E. coli* cells by fluorescent microscopy. Measurements were taken for various doses of IPTG treatment, after the gene expression dynamics reached steady state. Preliminary analysis of the fluctuations demonstrated that the distribution of G1 (CFP) is best described by a Gaussian of the logarithmic (Figure 2.2A, left) and not the linear abundance of protein (Figure 2.2A, right). Indeed, a Wilcoxon signed-rank test confirmed that the respective errors in the fit were not sampled from identical distributions ($p = 0.0013$), and the that the sum-squared errors was approximately four-fold lower for the log-normal fit. In addition, concomitant monitoring of G1 (CFP) and G2 (YFP) abundance for various doses of IPTG demonstrated that the joint distribution of gene products can be approximated by a bivariate Gaussian of the logarithmic protein abundance (Figure 2.2B-D). As a result, stochastic equations representing the dynamics of gene expression need to incorporate log-normally distribution fluctuations of protein abundance.

2.1.2 Noise in mammalian signal transduction

We characterized the fluctuations in phosphorylated protein abundances in the Mitogen Activated Protein Kinase (MAPK) cascade. The canonical MAPK cascade consists of a series of biochemical reactions, many of which involve the transfer of a phosphate group from ATP to specific amino acid residues of the down stream target protein by its parent kinase. The cascade involves the initial activation of MAPKKK, which processively phosphorylates MAPKK two times, which in turn, processively adds two phosphates to MAPK [39], (Figure 2.3). We measured the progression of signal through the MAPK pathway using monoclonal antibodies

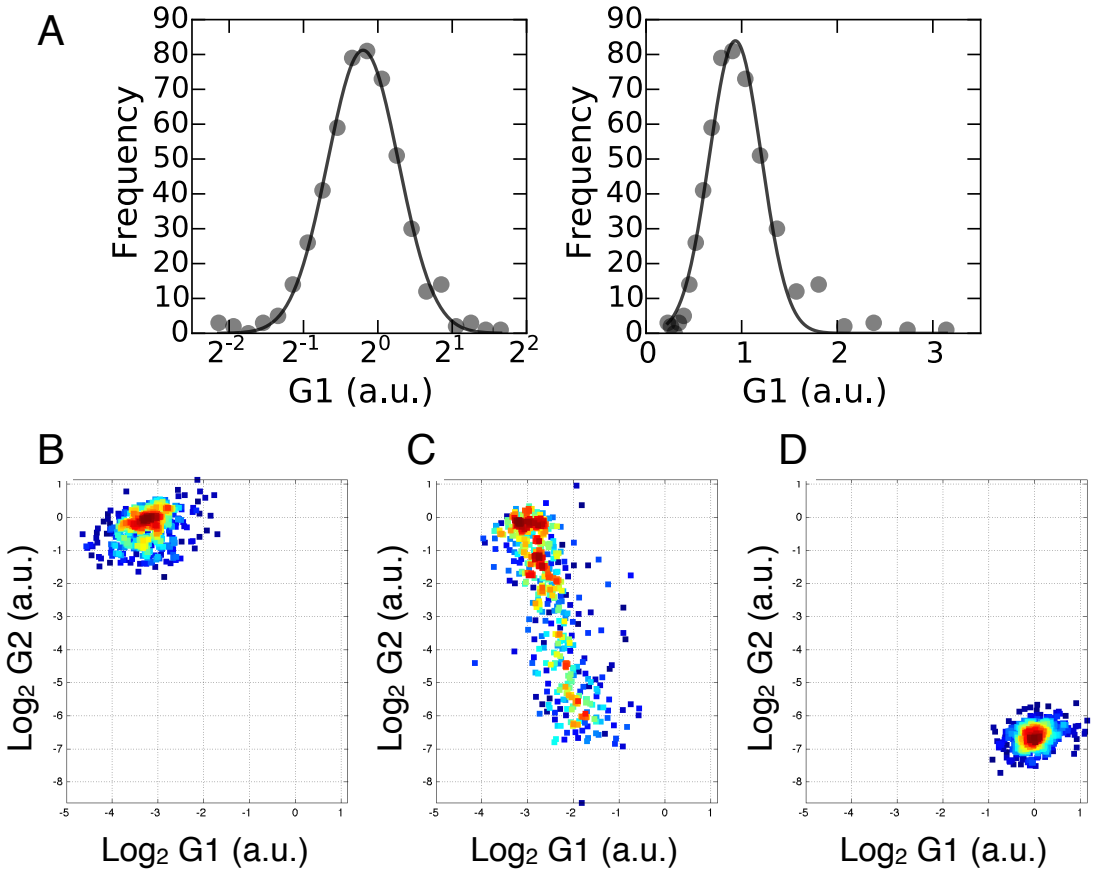

Figure 2.2: **Log-normally distributed protein in *E. coli*.** (A) Gaussian distribution fits to the distribution of G1 gene product (CFP) in log (right) and linear (left) scale. Single cell measurements of G1 (CFP) and G2 (YFP) for (B) a saturating dose, (C) intermediate dose, and (D) an ineffective dose of IPTG.

specific for the phosphorylated MEK 1/2 (p-Serine 221, MAPKK) and doubly phosphorylated ERK 1/2 (p-Threonine 202, p-Tyrosine 204, MAPK) in primary 5CC7 mouse $CD4^+$ T lymphocytes (T cells) by Flow Cytometry.

To stimulate the MAPK cascade in T cells we administered various doses of the small molecule phorbol 12-myristate 13-acetate (PMA). PMA activates the MAPK pathway by activating DAG-PKC-RasGRP signaling axis, which directly activates RAF followed by phosphorylating MEK (pMEK) and lastly phosphorylating ERK (ppERK) [24]. We measured the cells after ten minutes of stimulation, a sufficiently

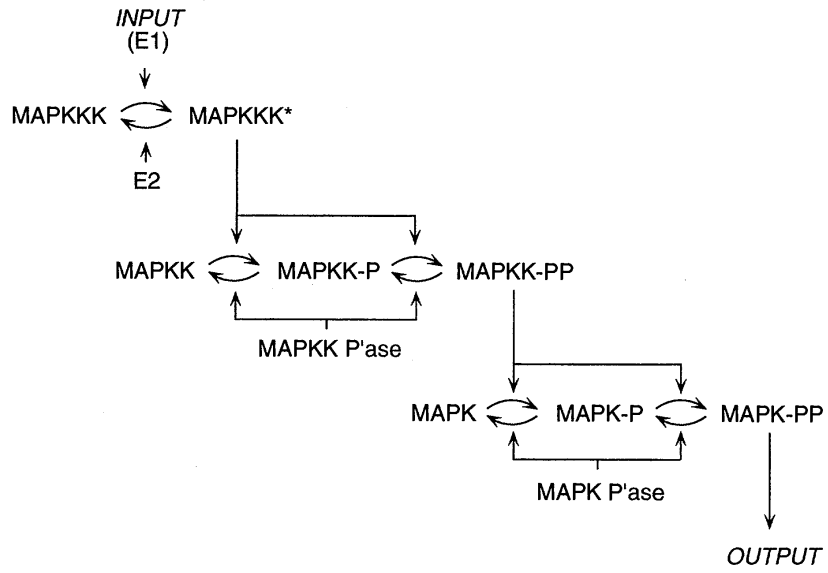

Figure 2.3: **Mitogen Activated Protein Kinase Cascade.** Reprinted with permission from PNAS: [39] copyright (1996). This model shows diagrammatically the organization of chemical reactions in the MAPK cascade.

long time for MAPK signaling to come to steady state [24].

Consistent with previous studies of the MAPK pathway we see that pMEK (Figure 2.4A) and ppERK (Figure 2.4B) are activated with increasing doses of PMA. Furthermore, analyzing the single cell response data (Figure 2.5) we see that pMEK and ppERK abundance per cell increases smoothly, sharply, and approximately according to a unimodal log-normal distribution. Quantitatively, the superiority of the log-normal distribution fit to the data over a normal distribution is determined by a statistical test and goodness of fit measure (Figure 2.6). Indeed, application of the Wilcoxon signed-rank test comparing the square residuals from each of the normal and log-normal distribution fits indicated the residuals were not equivalent ($p = 2.69 \text{ x } 10^{-14}$). In addition, the sum-square residuals showed approximately 10 fold preference to the log-normal description.

In summary, for both single cell data sets, gene expression in *E. coli* and

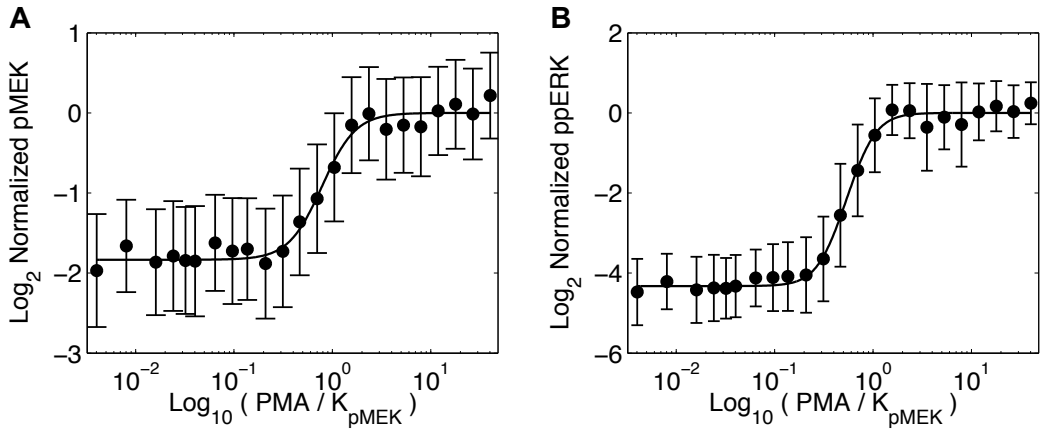

Figure 2.4: **PMA activation of T cells.** The average response of pMEK and ppERK in T cells activated with PMA. Error bars are standard deviation of the logarithmic abundance of pMEK and ppERK among individual cells.

phospho-response in T lymphocytes, we found that the distribution of biochemical species were log-normally distributed. In fact, this is a distribution that we find to be generic in most biological settings studied in the Altan-Bonnet Lab *i.e.* when measuring the abundance of surface markers, the secretion of cytokine by one individual cell, the rate of lymphocyte proliferation etc.

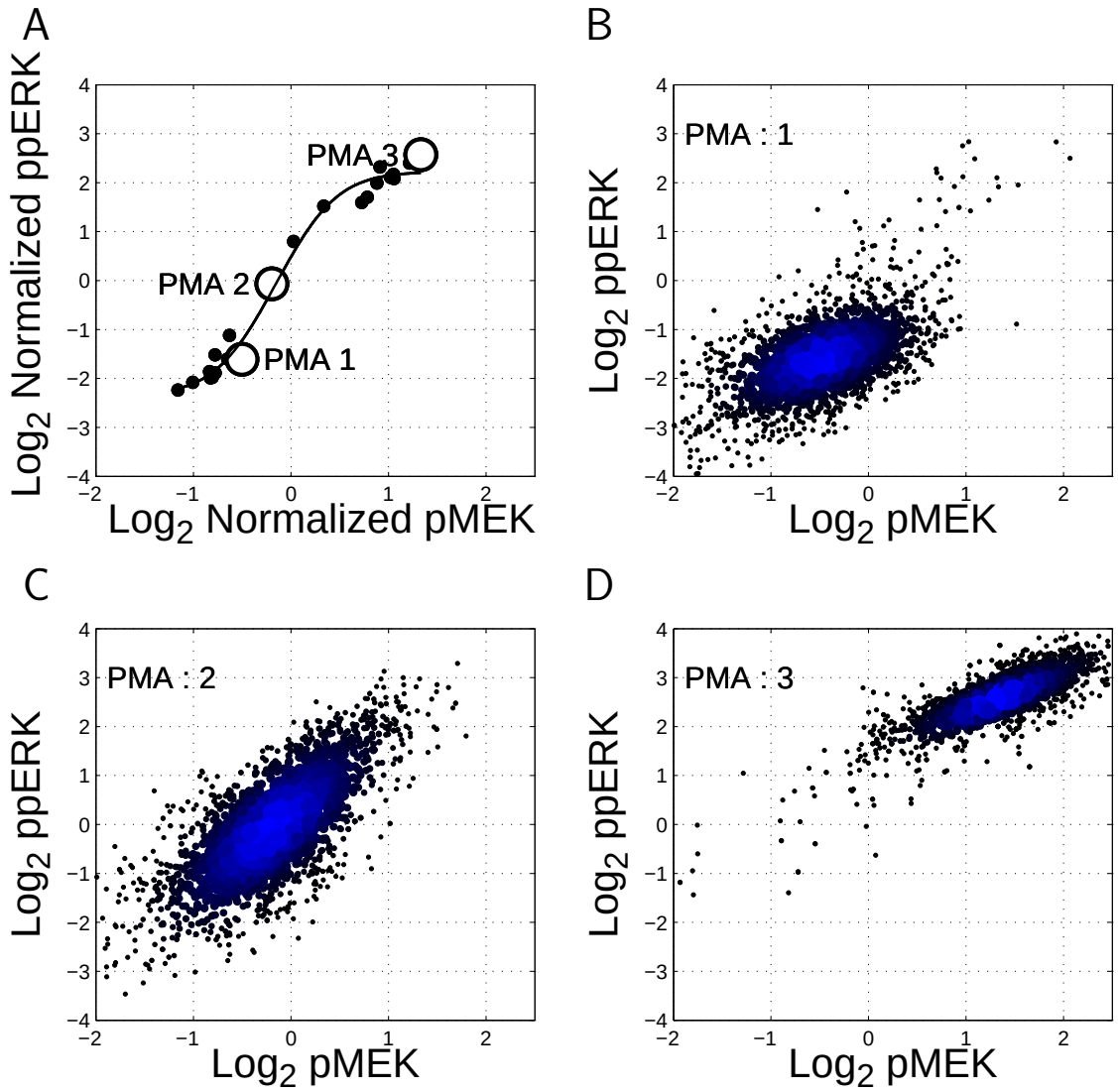

Figure 2.5: **Single cell pMEK and ppERK response to PMA.** (A) The response of ppERK to pMEK in PMA stimulated T lymphocytes. (B-D) Single cell flow cytometry measurements for the respective PMA dose. The increased size and shade of blue for a cell represents (pMEK, ppERK) abundances with higher probability. These figures demonstrate that pMEK and ppERK are activated unimodally and smoothly, qualitatively maintaining a bivariate log-normal distribution with increasing doses of PMA.

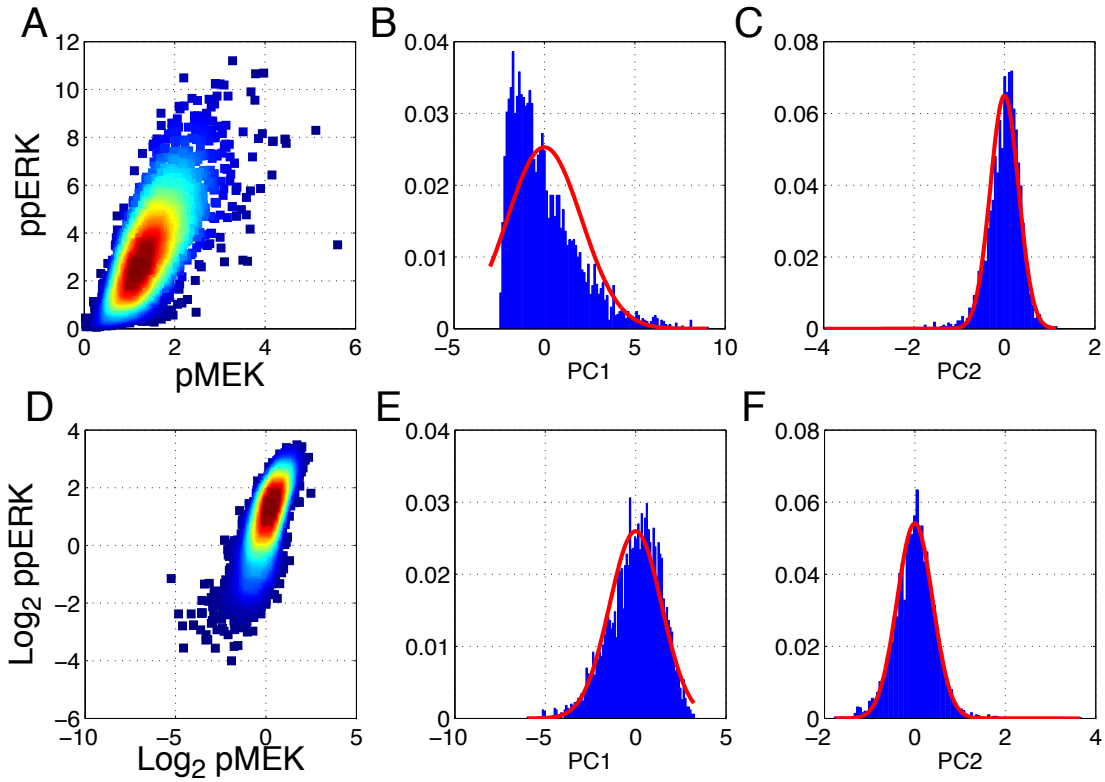

Figure 2.6: **Fluctuations in MAPK signaling are log-normally distributed.** Demonstration that the logarithmic representation of data is superior to the linear representation. (A) Bivariate distribution of the linear scale flow cytometry data. Gaussian distribution fits of the linear scale single cell data projected onto the first principle component (B) and the second principle component (C). (D) Bivariate distribution of the logarithmic scale single cell measurements. Gaussian distribution fits of the logarithmic scale single cell data projected onto the first principle component (E) and the second principle component (F).

2.2 Logarithmic chemical Langevin equation

We begin our stochastic treatment of biochemical reactions in single cells by writing the chemical Master equation as

$$\frac{d}{dt}p(n,t) = \Omega F p(n-1,t) + \lambda(n+1)p(n+1,t) - (\Omega F + \lambda n)\, p(n,t), \qquad (2.1)$$

where Ω is the volume of the cell, ΩF is independent of n and represents the probability of producing a single molecule per second, and lastly λ is a constant in which λn represents the probability of a single molecule degrading per second. Equation 2.1 represents the temporal evolution of the probability distribution of n particles observed in time. The goal of our analysis is to compute the magnitude of the fluctuations in the abundance of each molecular species per cell. As a result we only need to compute the mean and the variance of our system. With this in mind, we can approximate the chemical Master equation (Eq 2.1) to a simpler form that will provide the information we care about.

Consider that the volume of the cell is sufficiently large such that the production and degradation of individual molecules can be approximated by the continuous variable $x = \frac{n}{\Omega}$. Making this change of variables requires that,

$$p(n,t) = \int_{x=\frac{n-1/2}{\Omega}}^{\frac{n+1/2}{\Omega}} P(x,t)dx, \qquad (2.2)$$

taking $r = \frac{1}{\Omega}$ and substituting into Equation 2.1 we see that,

$$\frac{\partial P(x,t)}{\partial t} = \Omega F P(x-r,t) + \Omega\lambda(x+r)P(x+r,t) - \Omega(F+\lambda x)P(x,t), \qquad (2.3)$$

our new continuous variable chemical Master equation. Rewriting this equation with respect to van Kampen's translation operators [84],

$$W(x \pm r, t)P(x \pm r, t) = e^{\pm r\partial_x}\left[W(x,r)P(x,r)\right], \qquad (2.4)$$

we arrive at,

$$\frac{\partial P(x,t)}{\partial t} = \left[-\Omega(F+\lambda x) + e^{-r\partial_x}\Omega F + e^{r\partial x}\Omega\lambda x\right] P(x,t). \tag{2.5}$$

The advantage of this representations is that the entire equation is written in terms of $P(x,t)$. However, we still haven't simplified our equation. To do this we expand Equation 2.5 with respect to r about $r=0$,

$$\frac{\partial P(x,t)}{\partial t} = \sum_{n=1}^{\infty} \frac{r^n}{n!}\frac{\partial^n}{\partial x^n}\Omega\left((-1)^n F + \lambda x\right) P(x,t), \tag{2.6}$$

which is the Kramers Moyal Expansion of the Master equation. Indeed, we can now write an approximate solution of the chemical Master equation by truncating this series by assuming that r is small. We then truncate the series for $n > 2 \approx 0$,

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[F-\lambda x]P(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\frac{1}{\Omega}[F+\lambda x]P(x,t), \tag{2.7}$$

which is the Fokker-Planck equation of our system. The advantage of the Fokker-Planck description is that the chemical Master equation is now transcribed into a diffusion equation with drift. Consequently, the variance of the corresponding stochastic process exhibits fluctuations in the abundance of x as $\frac{1}{\Omega}[F+\lambda x]$ and the mean dynamics of x according to $F-\lambda x$. The resulting stochastic process can then be written as a Langevin equation,

$$\frac{dx}{dt} = F - \lambda x + \frac{1}{\Omega^{1/2}}\sqrt{F+\lambda x}\xi(t), \tag{2.8}$$

where $\xi(t)$ is a white noise random variable. We use this equation to describe the transmission of fluctuations in the abundance of molecules in biological networks. Indeed, we consider a set of biomolecular reactions (*e.g.* gene expression or phospho-signaling) whose dynamics are described by Equation 2.8, in which the production rate F is now dependent on the abundance of the interacting biomolecules,

$$\frac{dx_j}{dt} = \sum_{i\neq j}(f_j(x_i;\theta_{ji})) - \lambda_j x_j + \frac{1}{\Omega^{1/2}}\sqrt{\sum_{i\neq j}(f_j(x_i;\theta_{ji})) + \lambda_j x_j}\ \xi_j(t) \tag{2.9}$$

where $f_j(x_i;\theta_{ji})$ is the rate of production of biochemical species j controlled by biochemical species i, θ_{ji} is the set of parameters that determine the function $f_j(x_i;\theta_{ji})$, and λ_j is the rate constant for the degradation of species j. The production rate $f_j(x_i;\theta_{ji})$ is a nonlinear, saturating function representing enzymatic activity—*e.g.* the Hill equation. Writing the parameters as quantities that fluctuate around their mean values, $\lambda_j = \langle\lambda_j\rangle + \delta\lambda_j$, and $\theta_{ji} = \langle\theta_{ji}\rangle + \delta\theta_{ji}$, we have

$$\begin{aligned}\frac{dx_j}{dt} &= \sum_{i\neq j}\left(f_j(x_i;\langle\theta_{ji}\rangle + \delta\theta_{ji})\right) - (\langle\lambda_j\rangle + \delta\lambda_j)x_j \\ &\quad + \frac{1}{\Omega^{1/2}}\sqrt{\sum_{i\neq j}\left(f_j(x_i;\langle\theta_{ji}\rangle + \delta\theta_{ji})\right) + (\langle\lambda_j\rangle + \delta\lambda_j)\,x_j}\;\xi_j(t).\end{aligned} \tag{2.10}$$

The inclusion of the parameter fluctuations in the square root terms assumes that the intrinsic fluctuations $\xi_j(t)$ are faster than the parameter fluctuations $\delta\theta_{ji}$ and $\delta\lambda_j$. Writing x_j as a quantity that fluctuates around its mean value ($\langle x_j\rangle + \delta x_j$) we expand Equation 2.10 using the size of the fluctuations as the order parameter, resulting in terms of order zero (those that do not depend on the fluctuating terms), those that depend on the fluctuating terms to the first order, and those that depend on products of two or more fluctuating terms. Disregarding the latter, we obtain

$$\begin{aligned}\frac{d\langle x_j\rangle}{dt} + \frac{d\delta x_j}{dt} &= \sum_{i\neq j}\left(f_j(\langle x_i\rangle;\langle\theta_{ji}\rangle) + \frac{\partial f_j}{\partial x_i}\delta x_i + \frac{\partial f_j}{\partial\theta_{ji}}\delta\theta_{ji}\right) - \\ &\quad (\langle\lambda_j\rangle + \delta\lambda_j)\langle x_j\rangle - \langle\lambda_j\rangle\delta x_j \\ &\quad + \frac{1}{\Omega^{1/2}}\sqrt{\sum_{i\neq j}\left(f_j(\langle x_i\rangle,\langle\theta_{ji}\rangle)\right) + \langle\lambda_j\rangle\langle x_j\rangle}\;\xi_j(t)\end{aligned} \tag{2.11}$$

This equation can be solved hierarchically, order by order. In the stationary state, the zero-th order terms result in

$$\langle x_j\rangle = \frac{1}{\langle\lambda_j\rangle}\sum_{i\neq j} f_j(\langle x_i\rangle;\langle\theta_{ji}\rangle) \tag{2.12}$$

The first order terms give the equation for the fluctuations in species j

$$\frac{d\delta x_j}{dt} = \sum_{i \neq j} \left(\frac{\partial f_j}{\partial x_i} \delta x_i + \frac{\partial f_j}{\partial \theta_{ji}} \delta \theta_{ji} \right) - \langle \lambda_j \rangle \delta x_j - \delta \lambda_j \langle x_j \rangle + \frac{1}{\Omega^{1/2}} \sqrt{2 \langle \lambda_j \rangle \langle x_j \rangle}\, \xi_j(t) \tag{2.13}$$

Changing to logarithmic variables, let $y_j = \ln(x_j) = \ln(\langle x_j \rangle + \delta x_j)$. Then to the first order in δx_j we have

$$y_j = \ln(\langle x_j \rangle) + \frac{\delta x_j}{\langle x_j \rangle} + \mathcal{O}(\delta x_j^2) \tag{2.14}$$

and $\langle y_j \rangle = \ln(\langle x_j \rangle)$ given that $\langle \delta x_j \rangle = 0$. The logarithmic increment in species j is

$$\delta y_j = \ln(x_j) - \ln(\langle x_j \rangle) = \frac{\delta x_j}{\langle x_j \rangle} \tag{2.15}$$

Dividing Equation 2.13 by $\langle x_j \rangle$ we find the equations for the fluctuations in logarithmic scale

$$\frac{d\delta y_j}{dt} = \sum_{i \neq j} \left(\langle \lambda_j \rangle \frac{\partial \ln f_j}{\partial \ln x_i} \delta y_i + \frac{\langle \lambda_j \rangle}{\langle \theta_{ji} \rangle} \frac{\partial \ln f_j}{\partial \ln \theta_{ji}} \delta \theta_{ji} \right) - \langle \lambda_j \rangle \delta y_j - \delta \lambda_j + \sqrt{2 \frac{\langle \lambda_j \rangle}{\Omega \langle x_j \rangle}}\, \xi_j(t) \tag{2.16}$$

Equation 2.16 is the chemical Langevin equation with fluctuating parameters in log-scale variables linearized about a fixed point at $\langle x_j \rangle$.

When the abundance of the chemical species represented by $\Omega \langle x_j \rangle$ is large, the fluctuations represented by the last term in Equation 2.16 are small. Formally this is attained by making the volume Ω tend to infinity. Then, the only remaining fluctuations in Equation 2.16 are the fluctuations in the parameters, which we model as Wiener processes (i.e., fluctuations are fast compared with the half life of the species)

$$\sum_{i \neq j} \left(\frac{\langle \lambda_j \rangle}{\langle \theta_{ji} \rangle} \frac{\partial \ln f_j}{\partial \ln \theta_{ji}} \delta \theta_{ji} \right) - \delta \lambda_j = q_j \epsilon_j + q_G \epsilon_G \tag{2.17}$$

where the ϵ's are delta-correlated white noise processes. Approximating the left hand side by the sum of two random variables with constant coefficients is consistent with our experimental observations (discussed further in a proceeding section). Note that the parameter fluctuations are modeled by one term that is independent for each species ($q_j\epsilon_j$) and one term that is common to all species ($q_G\epsilon_G$) representing a global source of parameter fluctuations. Therefore, we have the linearized system

$$\frac{d\delta y_j}{dt} = \sum_{i \neq j} (a_{ji}\delta y_i) - \langle \lambda_j \rangle \delta y_j + q_j\epsilon_j + q_G\epsilon_G \tag{2.18}$$

where $a_{ji} = \langle \lambda_j \rangle \frac{\partial \ln f_j(\langle x_i \rangle, \langle \theta_{ji} \rangle)}{\partial \ln \langle x_i \rangle}$. The general form of these equations for arbitrary regulatory topologies $\mathbf{A}$ and normally distributed noise models $\mathbf{Q}$ is

$$\frac{d}{dt}\begin{bmatrix} \delta y_1(t) \\ \vdots \\ \delta y_N(t) \end{bmatrix} = \mathbf{A}\begin{bmatrix} \delta y_1(t) \\ \vdots \\ \delta y_N(t) \end{bmatrix} + \mathbf{Q}\begin{bmatrix} \epsilon_1(t) \\ \vdots \\ \epsilon_N(t) \\ \epsilon_G(t) \end{bmatrix} \tag{2.19}$$

where δy_i is the logarithmic increment in species i, the entries in the matrix $\mathbf{A}$ are a_{ji} as described above for $i \neq j$ and $a_{jj} = -\lambda_j$, and the entries of $\mathbf{Q}$ are independent of the log-concentrations.

In summary, log-scale fluctuations in the concentrations of reacting chemical species of sufficiently high concentration follow Equation 2.19 therefore the analysis presented in the main text applies to fluctuations in log-concentration of proteins.

2.3 Tracking the noise

Our premise is that the measured fluctuations of network elements is related to their connectivity. As a result the covariances and variances of measured observ-

ables among an ensemble of clonal single cells is reflective of the their interconnectivity. In order to determine the functional relationships between the measured observable covariances, we analyze a linear stochastic dynamical system where each species is represented by the stochastic differential equation

$$\frac{d}{dt}y_j(t) = \sum_{i=1}^{N} a_{ji}y_i(t) + q_{ji}\xi_i(t). \tag{2.20}$$

Here $y_j(t)$ represents the logarithmic fluctuations of biochemical species j from its mean abundance (for reading convenience we replace δy_j from the previous section with y_j), a_{ji} is the rate of change of species j in response to the presence of species i, q_{ji} is the contribution of the i^{th} species intrinsic noise to the j^{th} species intrinsic noise, and $\xi_j(t)$ is a white noise random variable. The white noise is characterized by two observations

$$\langle \xi_j(t) \rangle = 0 \text{ and}$$
$$\langle \xi_j(t)\xi_i(t+\tau) \rangle = \delta_{ji}\delta(\tau),$$

where δ_{ji} is the Kronecker delta function and $\delta(\tau)$ is the Dirac delta function.

The entire system of equations can be compactly written as

$$\frac{d}{dt}\mathbf{y} = \mathbf{A}\mathbf{y} + \mathbf{Q}\xi(t), \tag{2.21}$$

where the matrix $\mathbf{A}$ (dimension N x N) contains all of the network connectivity information and $\mathbf{Q}$ is an N x M matrix representing the magnitude of M noise sources to the N biochemical reactions. Note, that for theoretically ideal systems where the fluctuations in the dynamics are exclusively attributable to the reactions of each chemical species, $\mathbf{Q}$ would be an N x N diagonal matrix. Finally, $\mathbf{y}$ is an N dimensional column vector representing the fluctuations of each biochemical species and $\xi(t)$ is an M dimensional random white noise random variable. We proceed

by computing the correlation functions associated with the presented stochastic dynamical system.

The solution of the stochastic dynamical systems can be written as

$$\mathbf{y}(t) = \int_{s=-\infty}^{t} e^{\mathbf{A}(t-s)}\mathbf{Q}\xi(s)ds. \tag{2.22}$$

Here we see that the network connectivity described in $\mathbf{A}$ and the noise matrix $\mathbf{Q}$ determine the propagation of noise throughout the network. To obtain an explicit relationship between covariates, we compute the pairwise correlation functions between biochemical species. The pairwise correlation function takes the form

$$\mathbf{\Sigma}(\tau) = \left\langle \mathbf{y}(t)\mathbf{y}^T(t+\tau) \right\rangle, \tag{2.23}$$

$$= \int_{s'=-\infty}^{t+\tau} \int_{s=-\infty}^{t} e^{\mathbf{A}(t-s)}\mathbf{Q} \left\langle \xi(s)\xi(s')^T \right\rangle \mathbf{Q}^T e^{\mathbf{A}^T(t+\tau-s')} ds ds'. \tag{2.24}$$

Application of the properties of the white noise random variables $\xi(t)$,

$$\mathbf{\Sigma}(\tau) = \underbrace{\int_{s=-\infty}^{t} e^{\mathbf{A}(t-s)}\mathbf{Q}\mathbf{Q}^T e^{\mathbf{A}^T(t-s)} ds}_{\mathbf{\Sigma}_o} \; e^{\mathbf{A}^T\tau}, \text{ resulting in} \tag{2.25}$$

$$\mathbf{\Sigma}(\tau) = \mathbf{\Sigma}_o e^{\mathbf{A}^T\tau}, \tag{2.26}$$

where $\mathbf{\Sigma}_o$ is the unlagged covariance matrix (independent of τ), and $e^{\mathbf{A}^T\tau}$ represents the memory of the system to the random perturbations by ξ. In motivating the problem statement, we indicated that the average dynamics of our system is at steady state. We continue with the conjecture that the correlation function is also at steady state - a condition of stochastic processes referred to as weakly stationary. Invoking this condition by setting the time derivative of the correlation function to zero we see that

$$\frac{d}{dt}\mathbf{\Sigma}(\tau) = 0, \tag{2.27}$$

$$\mathbf{A}\mathbf{\Sigma}(\tau) + \mathbf{\Sigma}(\tau)\mathbf{A}^T = -\mathbf{Q}\mathbf{Q}^T e^{\mathbf{A}^T\tau} \tag{2.28}$$

Ideally, our task at this point would be complete. We have obtained explicit relationships between the network connectivity $\mathbf{A}$, noise matrix $\mathbf{Q}$, and the measurable correlation matrix $\mathbf{\Sigma}$. However, this representation requires measurements from a single cell sampled at a time interval much shorter than the reaction network correlation timescale. With this type of data, the temporal auto and cross correlation functions can be computed and compared to the theoretical predictions given the network connectivity and noise matrix $\mathbf{Q}$. However, our single cell measurements preclude any repeat measurements of the abundance of the biochemical species from the same cell.

We proceed by invoking the assumption of ergodicity - the equivalence of the ensemble statistics to the statistics of an individual entity measured for all time. While seemingly extreme, we have no reason to not assume ergodicity. The cells whose response we measure are clonal, grown identically, and perturbed in a well mixed homogeneous environment. Therefore, we continue with only the description of the measurable unlagged, $\tau = 0$, covariance written as

$$\mathbf{A}\mathbf{\Sigma}_o + \mathbf{\Sigma}_o\mathbf{A}^T = -\mathbf{Q}\mathbf{Q}^T. \tag{2.29}$$

Equation 2.29 is as a Lyapunov equation. The equation is used in dynamical systems to find network connectivities that result in stable dynamic solutions [51]. In contrast, in our system the existence of a stable solution is not in question, we exclude systems that are feedback regulated from this analysis. However, we are interested in how the network connectivity manifest into unique functional relationships between covariance elements, a topic we explore further in the next section.

2.4 Functional relationships between covariance elements

The key insight from the previous section is that the covariance matrix changes in a theoretically predictable way (specified by Equation 2.29) in response to external perturbations. The perturbations modulate the connectivity strengths between pairs of interacting components according to the physical constraints of the system studied. While the subsequent analysis is general and applicable to any dynamical system treated by linear Langevin equations, we focus specifically on the application to our biochemical systems (Eq 2.19).

We chose to describe the biological reactions in cellular signaling networks as a hill equation. The hill equation is a nonlinear function that describes the production of a biochemical product j in response to the abundance of biochemical species i, written as

$$f_{ji}(x_i;\theta_{ji}) = \nu_{ji}\frac{x_i^{n_{ji}}}{x_i^{n_{ji}} + \kappa_{ji}^{n_{ji}}} \qquad \text{for} \qquad \theta_{ji} = [\nu_{ji}, n_{ji}, \kappa_{ji}], \tag{2.30}$$

where x_i is the abundance of the biochemical species i. n_{ji} determines whether the biochemical reaction is inhibitory or activating, and controls the steepness of the interaction. Examining the possible values of n_{ji} we see that

$$n_{ji} \begin{cases} > 0 & \text{for activating interactions,} \\ < 0 & \text{for inhibiting interactions,} \\ = 0 & \text{for no interactions.} \end{cases}$$

The parameter ν_{ji} represents the amplitude of the response, and κ_{ji} represents the half effective abundance of i required to change the abundance of j by 50% of its maximal response. We chose this model for the following reasons: the resulting sigmoidal behavior of this function represents a good approximation to the behavior

of enzymatic reactions, and it is transparent - it is easy to ascertain the behavior of the function to changes in variables and parameters.

As a result of the nonlinear functional form of the Hill equation, we approximate the connectivity strengths, a_{ji}, to first order with respect to the logarithmic changes in x_i (Eqs 2.15,2.16,2.18),

$$a_{ji} = \langle \lambda \rangle \, k_{ji} = \langle \lambda \rangle \left. \frac{\partial \ln f_{ji}}{\partial \ln x_i} \right|_{x_i = \ln(\langle x_i \rangle)}, \tag{2.31}$$

in accordance with the linear Langevin systems theory.

In the subsequent sections we apply our Lyapunov equation (Eq 2.29) to measure the propagation of protein abundance fluctuations for both simulated and experimental data. Our simulated data are used to examine the unique propagation of fluctuations through signaling networks with different connectivities. We use our Lyapunov equation (Eq 2.29) to derive functional relationships between covariance elements with respect to the network configuration $\mathbf{A}$. We then proceed by applying our analysis to experimentally measured covariance elements, and attempt to extract information about the connectivity of the measured components. This analysis is performed in two experimental systems: 1) a synthetic *E. coli* gene regulatory system in which the data were collected and previously published by Pedraza and van Oudenaarden [64] and 2) MAPK phospho-signaling in primary T lymphocytes.

2.4.1 Noise propagation in the linear two node network

Consider a network composed of a single directional edge spanning two components, y_1 and y_2 (Figure 2.7A), and a diagonal noise matrix,

$$\frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} -\lambda & 0 \\ a_{21} & -\lambda \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \underbrace{\begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix}. \tag{2.32}$$

Where, λ is the rate characterizing the decay of each components abundance, and a_{21} represents the strength of the interaction between biochemical species 1 and 2 as defined Equation 2.31. The corresponding solution to the Lyapunov equation (Eq 2.29) for the covariance elements with respect to $\mathbf{A}$ and $\mathbf{Q}$ provides our desired description of the transmitted fluctuations. The resulting covariance matrix elements are:

$$\sigma_1^2 = \alpha, \tag{2.33a}$$

$$\sigma_{12} = \frac{\alpha a_{21}}{2\lambda}, \text{ and} \tag{2.33b}$$

$$\sigma_2^2 = \beta + \frac{\alpha a_{21}^2}{2\lambda^2}; \tag{2.33c}$$

where $\alpha = \frac{q_1^2}{2\lambda}$ and $\beta = \frac{q_2^2}{2\lambda}$ represent the contribution of the intrinsic fluctuations to the respective variances.

By examining the expressions representing the variances of y_1 and y_2, we clearly see the unidirectional transmission of fluctuations from the the first component (y_1) to the second component (y_2). Indeed, the variance of y_1 is exclusively dependent on the intrinsic biological fluctuations, α. In contrast, the variance of y_2 depends on both its intrinsic biological fluctuations β and the those transmitted from the first component α. The extent that fluctuations are propagated from the first to the second component is tuned by the connectivity strength, the stronger the connection the higher the propagation.

Indeed, experimental tools to modify the connectivity strength are often limited to exogenously administrated activating signal. Analogously, we performed an *in silico* dose response and examined the connectivity strength for changes in $\ln(\langle x_1 \rangle)$, represented by $\tilde{y}_1$, abundance. We see that a_{21} non-monotonically changes with the increasing abundance of $\tilde{y}_1$ - first a_{21} increases, reaches a maximum, and then decays (Figures 2.7B,C), an obvious consequence to taking the derivative of a sigmoidal function. We then examined how the fluctuations from y_1 propagate to y_2 by evaluating σ_2^2 (Equation 2.33) for each dose of activating signal.

Instead of explicitly evaluating the dependence of σ_2^2 with respect to a_{21} - a quantity often difficult to experimentally measure - we propose an alternative approach that complements our single cell experiments. Our approach is to rely upon the theoretical proportionality of covariance elements between y_1 and y_2 and the connectivity strengths. By reexamining our solutions in Equations 2.33 with respect to the variances of y_1 and y_2 and the covariance σ_{12} instead of a_{21}, we see that

$$\sigma_1^2 = \alpha, \text{ and} \tag{2.34a}$$

$$\sigma_2^2 = \beta + 2\frac{\sigma_{12}^2}{\alpha}. \tag{2.34b}$$

The theory now emphasizes that noise propagates in the system as the quadratic of the covariance between interacting components. In accordance with these equations, our *in silico* experiments recover the invariance of σ_1^2 to σ_{12} (Figure 2.7D), and the quadratic dependence of σ_2^2 on σ_{12} (Figures 2.7E). Here, we have theoretically demonstrated how the functional relationships of two very simple experimental measurements can track the fluctuations from one biochemical species to another.

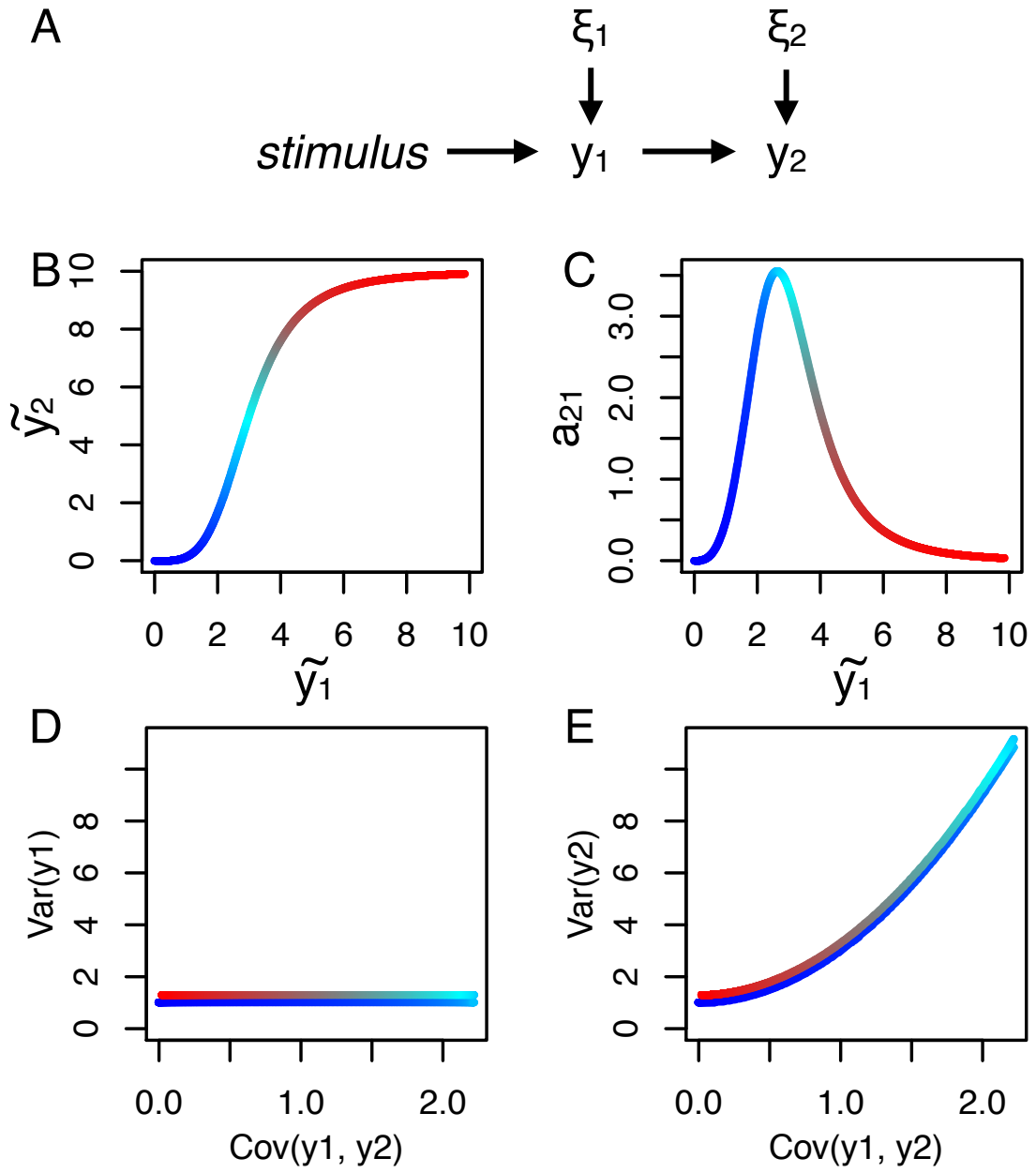

Figure 2.7: **Propagation of fluctuations in the two node system.** (A) Diagram of the simple two node network. (B) The response of $\ln \langle x_2 \rangle = \tilde{y}_2$ with respect to changes of $\ln \langle x_1 \rangle = \tilde{y}_1$. The color map from blue to red indexes the stimulation strength. (B) Changes in the connectivity strength in response to our *in silico* titration experiments. (C) The corresponding response of the the variance of y_1 with respect to the covariance of y_2. (D) The quadratic dependence of the variance of y_2 with respect to the covariance σ_{12}.

2.4.2 Noise propagation in the fan-in network

Next we investigated the functional relationships of covariance elements of the fan-in network. In the fan-in network, component two (y_2) is regulated by components one (y_1) and three (y_3, Figure 2.8A), and is described by the following dynamical system,

$$\frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underbrace{\begin{bmatrix} -\lambda & 0 & 0 \\ a_{21} & -\lambda & a_{23} \\ 0 & 0 & -\lambda \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \underbrace{\begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}. \tag{2.35}$$

Solving for the covariance matrix elements according to the Lyapunov equation, (Equation 2.29) the network connectivity, $\mathbf{A}$ and noise $\mathbf{Q}$ yields:

$$\sigma_1^2 = \alpha, \tag{2.36a}$$

$$\sigma_3^2 = \gamma, \tag{2.36b}$$

$$\sigma_{12} = \frac{\alpha a_{21}}{2\lambda}, \tag{2.36c}$$

$$\sigma_{13} = 0, \tag{2.36d}$$

$$\sigma_{23} = \frac{\gamma a_{23}}{2\lambda}, \text{ and} \tag{2.36e}$$

$$\sigma_2^2 = \beta + 2\frac{\sigma_{12}^2}{\alpha} + 2\frac{\sigma_{23}^2}{\gamma}, \tag{2.36f}$$

where $\alpha = \frac{q_1^2}{2\lambda}$, $\beta = \frac{q_2^2}{2\lambda}$, and $\gamma = \frac{q_3^2}{2\lambda}$. As a result of y_1 and y_3 being independently regulated by the signal, the fluctuations from each respective source contributes independently to the variance of y_2 (Equation 2.36f). Consequently, the covariance elements in the fan in network may exhibit functional relationships that deviate from the ideal linear two node system (Figure 2.7E, 2.8E). The extent of the deviation depends on the parameterization of the model.

Figure 2.8, for example, is parameterized such that the *in silico* activation of

the network results in a (σ_{12},σ_2^2) cycle or loop. The loop is a manifestation of the staggered activity of component two with respect to the activity of components one and three (Figure 2.8). Indeed, the specific parameters are such that the sensitivity $\kappa_{23} < \kappa_{21}$ results in a_{23} reaching a higher maximum and at lower abundance of $\tilde{y}_3$ than that of a_{21} (Figure 2.8C). The activity of component one is only dependent on the stimulus strength and its intrinsic biological fluctuations, which results in a σ_1^2 invariant to σ_{12} (Figure 2.8D). Meanwhile, the staggered connectivities of component one and three contribute to the fluctuations of σ_2^2 for different stimulus activity. The loop is a combined result of the staggered connectivities of the upstream components.

The staggered connectivity strengths determine the order in which (σ_{12},σ_2^2) ordered pairs change with respect to increases of stimulation strength. Consider the following table (Table 2.1) that separates the stimulus activity into three simple regimes. When the stimulation strength is low, the effective network is composed of

Table 2.1: **Analysis of the fan in network.** Analyzing the system of equations in three simple regimes: low, medium, and high stimulating conditions emphasizes how fluctuations are transmitted through the network.

Stimulation Strength	Connectivity Strength	Covariance	Effective network
low	$a_{21} < a_{2,3}$	$\sigma_{12} < \sigma_{2,3}$	$y_3 \to y_2$
medium	$a_{21} = a_{2,3}$	$\sigma_{12} = \sigma_{2,3}$	$y_3, y_1 \to y_2$
high	$a_{21} > a_{2,3}$	$\sigma_{12} > \sigma_{2,3}$	$y_1 \to y_2$

only component three activating component two. The (σ_{12},σ_2^2) plane represents this as in an increase of σ_2^2 but not σ_{12}. Increasing the stimulation strength to medium will results in equal contributions of both components, where a_{21} crosses a_{23} in Figure 2.8C, leading to an increase in both the σ_{12} and σ_2^2. Lastly, by increasing the stimulation to high results in the effective network of component one activating component two. The corresponding behavior in the (σ_{12},σ_2^2) plane converges to the isolated two component quadratic dependence first shown in Figure 2.7E. This is

obvious from the model equations in the high dose regime; $\sigma_{23} \propto a_{23} \to 0$ (Eq 2.36e) resulting in the simplification of Equation 2.36f into the two node model expression (Eq 2.34b). The overall result is a loop pattern shown in Figure 2.8E.

An attractive feature of tracking the noise in biochemical networks is the characterization of unmeasured components. Indeed, the presence of a cycle in the covariance and variance plane is informative of unmeasured quantities, however the converse is not true. The absence of a loop does not eliminate the possibility of unmeasured biochemical interactions. For example if $a_{21} = a_{23}$ for all doses, the data projected onto the covariance and variance plane would exhibit the two component quadratic behavior. Although, this limitation is restricted to semi-quantitative functional relationships discussed so far, it does not preclude the ability to infer unmeasured interactions from detailed quantitative assessments.

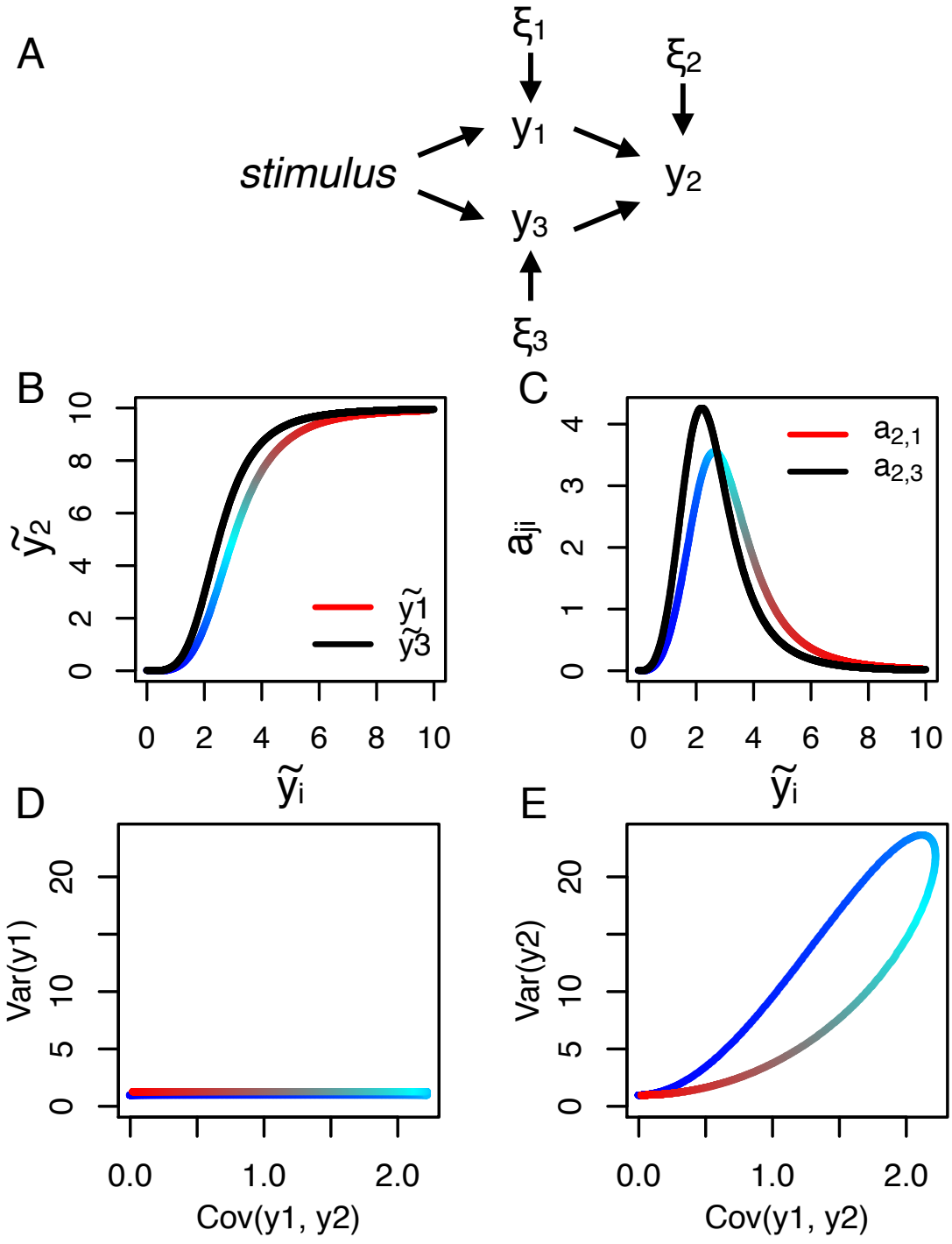

Figure 2.8: **Propagation of fluctuations in the fan in system.** (A) Diagram of the fan in network structure. (B) The responses of $\tilde{y}_2$ with respect to $\tilde{y}_1$ and $\tilde{y}_3$ exclusively, as before activating signal increases from blue to red. (C) The connectivity strengths of y_2 corresponding to either y_1 or y_3. (D) The variance of y_1 remains invariant to σ_{12}. (E) The (σ_{12},σ_2^2) plane shows a signature of fluctuations of y_2 transmitted by both y_1 and y_3.

2.4.3 Noise propagation in a cascade network

Lastly, we consider the cascade network, in which a stimulus activates component three, which activates component one, followed by component one activating component two. The dynamics of this network are described by

$$\frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underbrace{\begin{bmatrix} -\lambda & 0 & a_{13} \\ a_{21} & -\lambda & 0 \\ 0 & 0 & -\lambda \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \underbrace{\begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}. \tag{2.37}$$

Solving for the covariance elements in Equation 2.29 with respect to the cascade network yields:

$$\sigma_3^2 = \gamma, \tag{2.38a}$$

$$\sigma_{13} = \frac{\gamma a_{13}}{2\lambda}, \tag{2.38b}$$

$$\sigma_1^2 = \alpha + 2\frac{\sigma_{13}^2}{\gamma}, \tag{2.38c}$$

$$\sigma_{23} = \frac{\gamma a_{21} a_{13}}{4\lambda^2}, \tag{2.38d}$$

$$\sigma_{12} = \frac{\alpha a_{21}}{2\lambda} + \frac{3\gamma a_{21} a_{13}^2}{8\lambda^3}, \text{ and} \tag{2.38e}$$

$$\sigma_2^2 = \beta + 2\frac{\gamma\sigma_{12}^2}{\alpha\gamma + 3\sigma_{13}^2}, \tag{2.38f}$$

where $\alpha = \frac{q_1^2}{2\lambda}$, $\beta = \frac{q_2^2}{2\lambda}$, and $\gamma = \frac{q_3^2}{2\lambda}$. In this network the fluctuations of component one are now dependent on the stimulation strength by the connectivity a_{13} (Eq 2.38c). In addition, the measured σ_{12} is no longer a representation of component one and two exclusively, but depends also on the connectivity a_{13}. As a result, both σ_{12} and σ_1^2 are functionally dependent on the connectivity a_{13} which manifests as a loop in the (σ_{12},σ_1^2) plane (Figure 2.9D, we will explore alternative parameterizations later). Meanwhile the description of σ_2^2 is complex for two reasons: 1) σ_{12} is dependent on two connectivities, and 2) the propagation of fluctuations are

dependent on the quotient of σ_{12} and σ_{13}. Indeed, the cascade network is easily identifiable by the presence of loops for both covariance and variance planes in Figures 2.9D,E, however the precise shape of the possible (σ_{12},σ_2^2) representations are diverse (Figure 2.10).

We next investigated the sensitivity of our covariance and variance representation to the model parameters of the cascade network. Figure 2.10 shows our systematic change in the model parameters. Each row of Figure 2.10 corresponds to increases in both the maximum connectivity strength and the sensitivity of a_{21} (Figure 2.10B,F,J,N, column $a_{ji}(\tilde{y}_3)$). The sensitivity of a_{ji} represents the abundance of $\tilde{y}_3$ in which a_{ji} is maximum. From this analysis two features become apparent. First, the simulated data shows that the size of loops in the covariance and variance representations are dependent on the relative sensitivities. When the sensitivities of the a_{ji}'s are equivalent, we see the complete elimination of loops in the covariance and variance planes (Figures 2.10K,L). In contrast, the size of the loops increased with the increase separation of the connectivity sensitivities (Fig 2.10, columns $\sigma_1(\sigma_{12})$ and $\sigma_2(\sigma_{12})$). Secondly, we see that the relative sensitives of a_{ji}'s determine the order in which the increasing stimulation strengths change the respective covariances and variances. For example, consider Figure 2.10C and G, here the sensitivity of a_{13} is greater than a_{21} and results in an increase of σ_1^2 without a corresponding increase in σ_{12}. Indeed this result is a general feature of our covariance and variance representation. Increases of variance elements without the corresponding change in the covariance occurs when the covariance does not account for the fluctuations propagated to the measured variance, *e.g.* the system is more sensitive to an unmeasured quantity. Indeed, the opposite is true when the connectivity between measured pairs (y_1,y_2) is more sensitive than the contribution from the third unmeasured component (Figure 2.10O). In this example

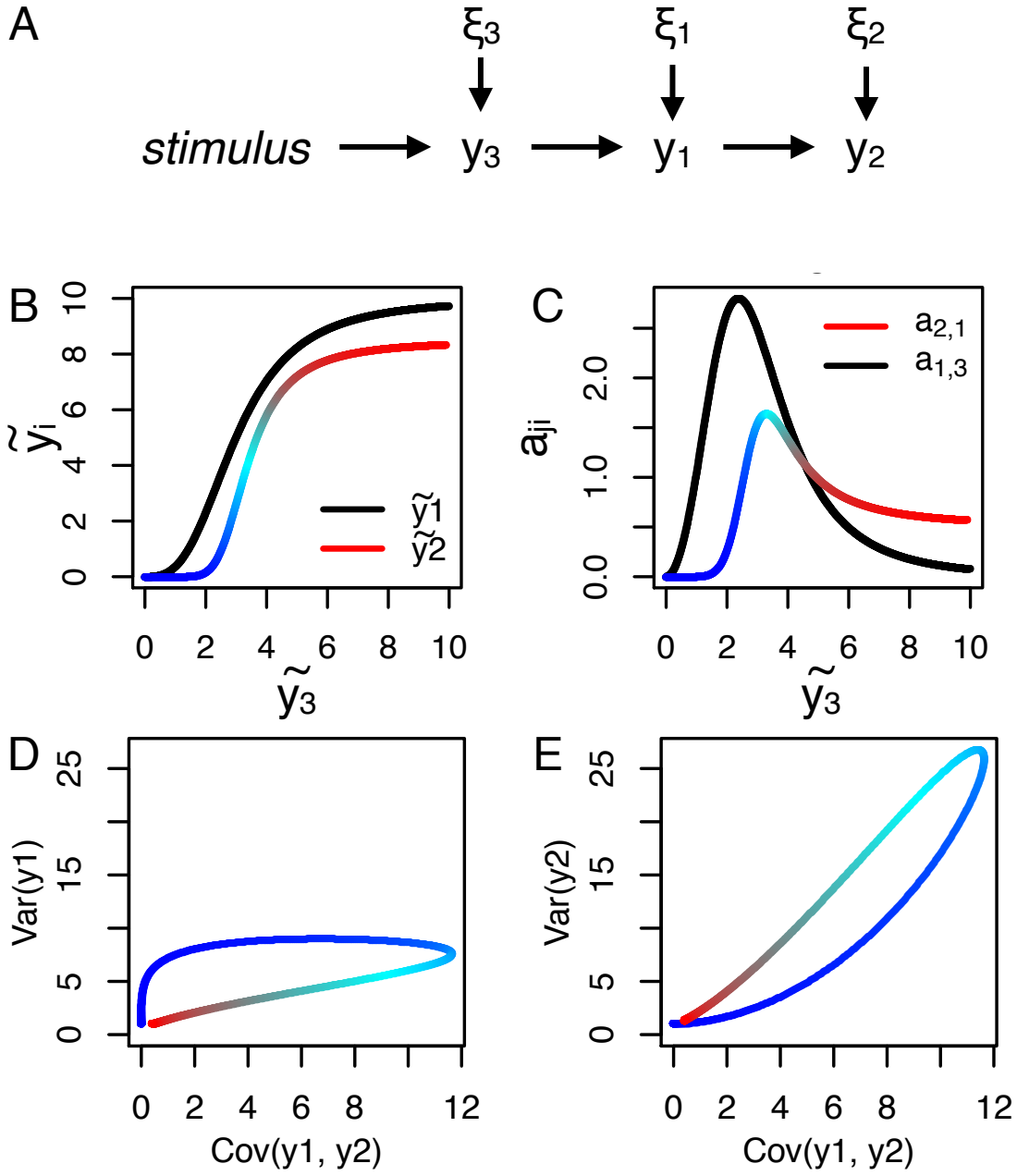

Figure 2.9: **Propagation of fluctuations through the cascade network.** (A) Diagram of the cascade network structure. (B) The responses of $\tilde{y}_1$ and $\tilde{y}_2$ to changes in $\tilde{y}_3$, for increasing stimulus strength (blue $\rightarrow$ red). (C) The connectivity strengths a_{21} and a_{13} representing each pair of biochemical reactions. (D) The covariance and variance plane of y_1. (E) The covariance and variance plane for y_2.

we first see a concomitant increase of both σ_{12} and σ_2^2, because the fluctuations of y_2 are accounted for by the covariance. As the stimulation strength increases the fluctuations from the unmeasured component increase while the measured decrease. As a result σ_2^2 increases despite the reduction in the measured σ_{12}. Taken together the propagation of fluctuations are highly dependent to the connectivity of the network.

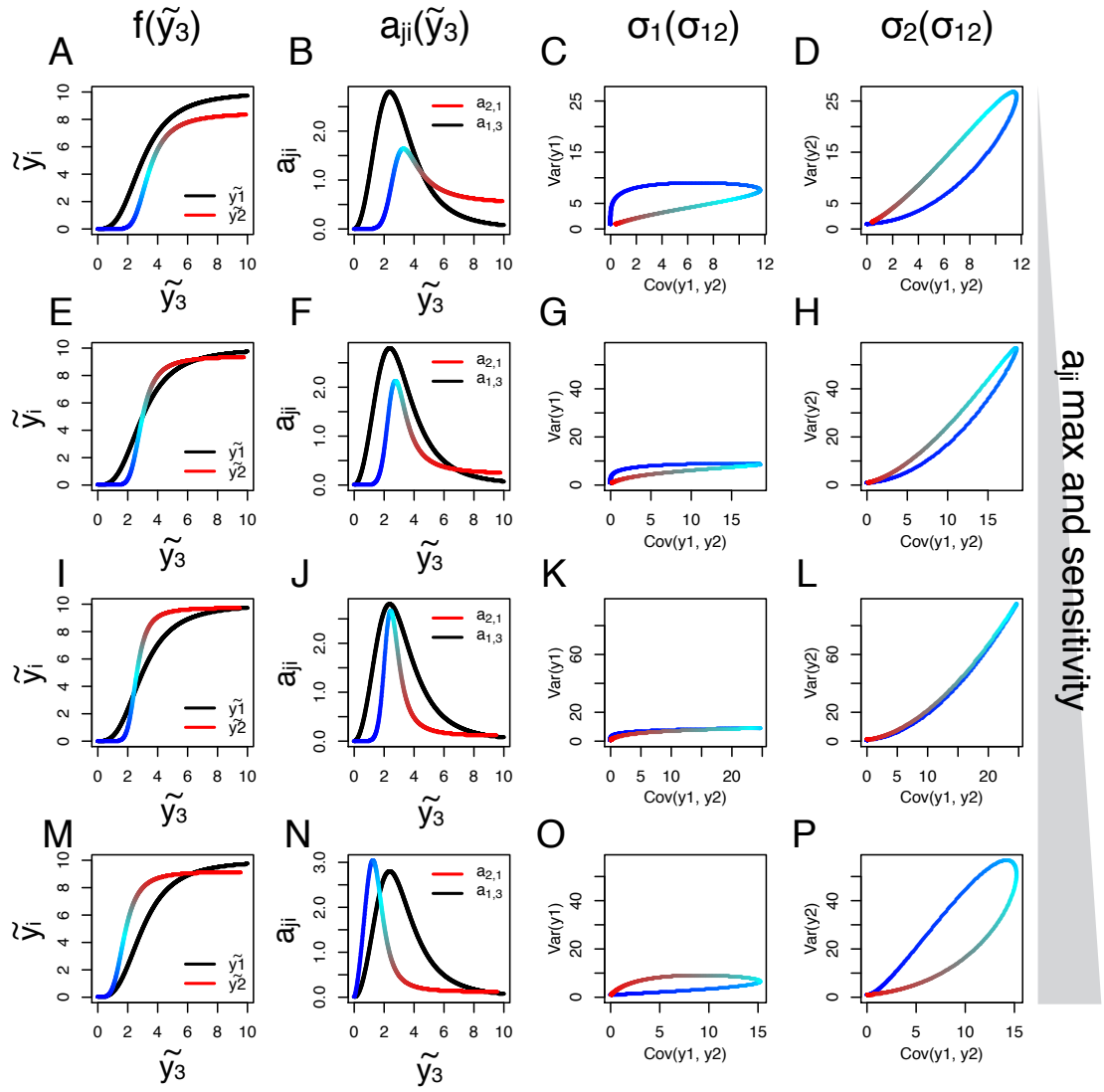

Figure 2.10: **Parameter sensitivity in the cascade network.** (A) The logarithmic responses of component one and two with increasing component three. (B) The connectivity strength for each value of $\tilde{y}_3$. (C) Characterization of the propagation of fluctuations measured in the (σ_{12},σ_1^2) plane. (D) Characterization of the propagation of fluctuations measured in the (σ_{12},σ_2^2) plane. The model presented in (A-D) is identical to that of Figure2.9, we include these graphs to facilitate comparisons to changes in model parameterization. (E-P) Each row represents the identical column-wise analysis of (A-D) but with unique parameterization of the model functions. The parameters change such that the amplitude of the maximum connectivity strength a_{21} and the sensitivity of a_{21} to $\tilde{y}_3$ increase. The sensitivity of a_{21} is the abundance of y_3 such that a_{21} is at maximum strength.

2.5 Noise propagation in single cells

To see if our analysis captures the properties of noise propagation in living systems we tested our theory on two known signaling systems. First, on published data of gene activity in a synthetic negative regulatory system expressed in *E. coli* [64]. Secondly, in Mitogen Activated Protein Kinase (MAPK) phospho-signaling in T lymphocytes. Each experimental model system consisted of single cell measurements that concomitantly quantified the abundance of each interacting biochemical species. We then estimated the covariance elements from the ensemble of single cell measurements for different doses of exogenously administered stimuli. With these data we were able test our the functional relationship of covariance elements in the isolated two component model.

Application of our theoretical predictions required that we account for experimental limitations. For each study we measured fluorescent markers that correlated with the abundance of our molecule of interest. However, these measurements are prone to additional noise introduced by the autofluorescence of cells, and / or non-specific binding of fluorescent molecules. We accommodated for these sources of noise with the extra terms (q_g, ξ_g) resulting in

$$\frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -\lambda & 0 \\ a_{21} & -\lambda \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} q_1 & 0 & q_g \\ 0 & q_2 & q_g \end{bmatrix}\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_g \end{bmatrix}, \tag{2.39}$$

dynamic equations. We then solved for the covariance elements using our derived Lyapunov equation (Eq 2.29), which yielded,

$$\sigma_1^2 = \alpha + \eta_g \text{ and} \tag{2.40a}$$

$$\sigma_2^2 = \beta + \eta_g\left(1 - 2\frac{\sigma_{12}}{\sigma_1^2}\right) + 2\frac{\sigma_{12}^2}{\sigma_1^2}. \tag{2.40b}$$

If we take $\eta_g \to 0$, we return to our theoretical results of the isolated two component system (Eq 2.34).

Lastly, we demonstrate how our model of the covariance elements facilitates the inference to the *direction* of the biochemical reactions. Equation 2.39 supposes that the abundance of y_2 changes in accordance to the dynamics of y_1. However, we could have easily built the opposite model. We can test which of these models represent the 'true' model by comparing the residuals between the fit of each model and our data. The estimated parameters $\hat{\theta}^m$ of model m were numerically found by minimizing the sum square residuals. The model with consistently smaller errors represents the most likely direction of the biochemical interaction. We statistically evaluated the best description of the data by performing two statistical tests. First, we used a two-sided non-parametric Wilcoxon singed rank test, whose null hypothesis is that the difference of paired residuals from each model are sampled according to a symmetric distribution around zero. Indeed, a rejection of this hypothesis indicates that one model out performs the other. If the data are sufficiently different we proceed by using a one-sided Wilcoxon signed-rank test to obtain a p-value representing our certainty of the best model.

2.5.1 Tracking fluctuations in an *E. coli* gene regulatory network

We analyzed the published data from Pedraza and van Oudenaarden [64] to test our noise propagation methodology. The experiment consisted of exogenous doses of IPTG and single cell measurements of steady state abundance of CFP (G_1) and YFP (G_2). By increasing the dose of IPTG they were able to strengthen the induc-

tion of g_1 (Figure 2.1) and consequently reduce the expression of g_2 (Figure 2.1). The concomitant measurements of G_1 and G_2 per cell allows us to correlate the average response of G_2 with respect to G_1 (Figure 2.1C). We see that the average response is well approximated by the hill equation (Eq 2.30), with $n < 0$ representing the inhibitory interaction. We next tested our covariance and variance analysis in characterizing the propagation of noise from $G_1$to G_2.

Figure 2.11D shows our estimation of covariance elements and the resulting isolated linear two component prediction. From our analysis we see two important features in our data. First, our estimation of σ_{G_1,G_2} is negative for IPTG doses in the dynamic range of the average G_1 and G_2 response. Indeed this result complements our developed theory, which states that the covariance between two elements is proportional to the connectivity strength (Eq 2.33b). We characterized the connectivity strength by the logarithmic gradient of the hill equation (Eq 2.31). The maximum absolute value of the log-gradient corresponds to the position of G_1 with induces the greatest response in G_2. Furthermore, the sign of the covariance originates from $n < 0$ and is indicative of the inhibitory interaction (Figure 2.11C). Secondly, the variance corresponding to both genes qualitatively behaves in accordance with our predictions. Namely, that $\sigma^2_{G_1}$ is invariant to the measured covariance, while $\sigma^2_{G_2}$ increases roughly by the square of the covariance (Fig 2.11D). Our model and the corresponding data captures our intuition, that noise propagates unidirectionally in biochemical reactions, however the deviation between the measured G_2 variance and model requires closer attention.

The dispersion of G_2's variance with respect to the isolated two node model is likely a consequence that the regulatory circuit is best represented as a cascade, and not our simplified two node model. Recall that G_1 is activated by G_0, and

that the dosage of IPTG inhibits G_0's negative regulation of G_1. Indeed, our analysis of the cascade network (Figure 2.10) showed how fluctuations from G_0 could have small influences to the measured variance of G_1 and generate a small loop for the variance of G_2. We did not pursue incorporating the more complex model to our data. This is because, the number of model parameters with this additional component increases our current number of parameters from two to five, and would likely be over-fitting our data. The over-fitting could be circumvented by measurements of active G_0 in addition to G_1 and G_2, however these measurements are not available.

We conclude our analysis of the *E. coli* gene data by applying our directionality test. Figure 2.12 shows the optimal fit of the isolated two node model for both network directions. Indeed, the true direction of the biochemical interaction is easily determined by inspection, and confirmed by our prescribed statistical analysis, $p = 1.2 \text{ x } 10^{-7}$ for $N = 23$ data points representing each IPTG dose.

In summary our theoretical description of the properties of noise propagation is in agreement with the data from the synthetic gene regulatory network first studied by Pedraza and van Oudenaarden [64].

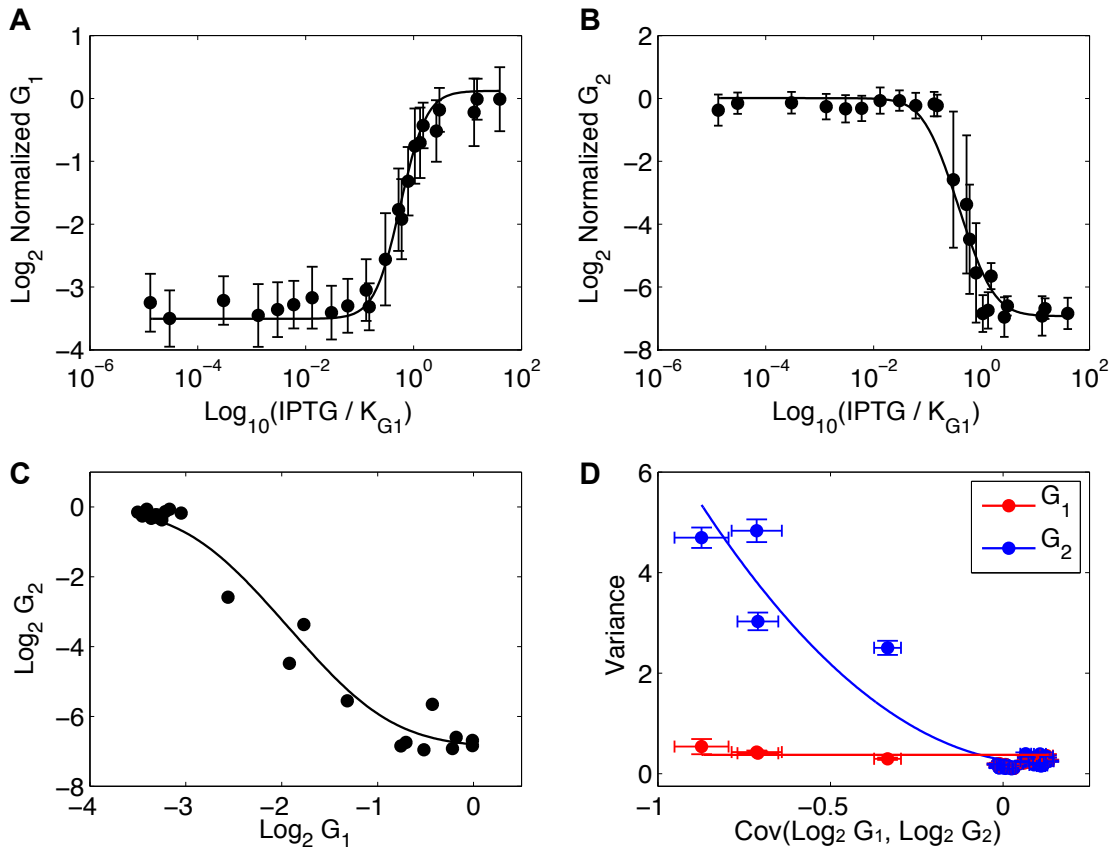

Figure 2.11: **Noise propagation in an *E. coli* gene regulatory network** (A) The logarithmic change of G_1 expression with increasing exogenous doses of IPTG. (B) The logarithmic change of G_2 expression with increasing exogenous doses of IPTG. Error bars in (A-B) represent the standard deviation of the respective gene expression among cells. (C) The average response of G_2 to changes in G_1. (D) The measured $(\sigma_{G_1,G_2},\sigma_i^2)$ for $i = \{G_1, G_2\}$ of single cell gene expression for various doses of IPTG. Error bars are $\pm$ one standard deviation computed from 1000 bootstraps. The respective lines for G_1 and G_2 represent the best fit of our model (Eq 2.40) from the linear two node configuration.

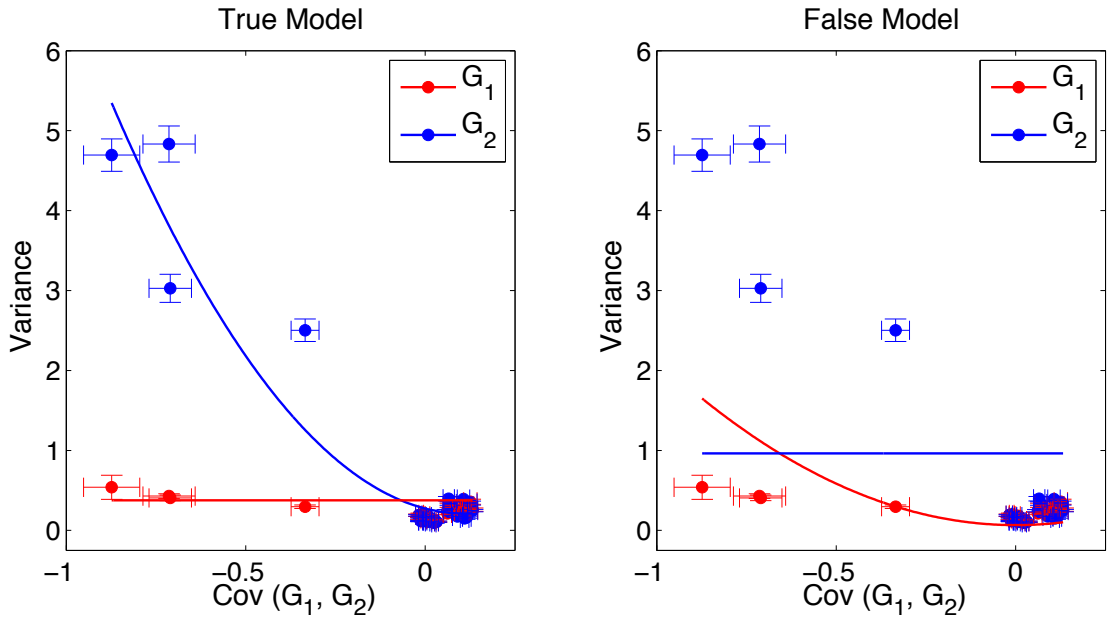

Figure 2.12: **Gene regulation directionality test** The covariance and variance plots with competing model fits. The true model represents a fit of covariance elements for the designed interaction, $G_1 \rightarrow G_2$. In contrast the false model fits the covariance elements according to the interaction $G_2 \rightarrow G_1$.

2.5.2 Tracking fluctuations in MAPK signaling of PMA activated T lymphocytes

We extended our theoretical description of noise propagation in biochemical systems to phospho-signaling in T lymphocytes. We measured the transmission of phosphates among two members of the Mitogen Activated Protein Kinase Cascade, namely phosphorylated MEK (pS221) and doubly phosphorylated ERK (pT202,pY204) by flow cytometry. To activate this signal transduction cascade and modulate the connectivity strengths between these pMEK and ppERK, we treated the cells with various doses of PMA, a potent activator of the MAPK pathway (Figures 2.13A,B). As before with the *E. coli* data, the concomitant measurements of pMEK and ppERK allow us to correlate the average response of ppERK to pMEK. These data were well approximated by the Hill equation (Eq 2.30) with $n > 0$. We next tested the ability of our theoretical description of the isolated two node system in describing the transmission of fluctuations from pMEK to ppERK.

Figure 2.13D shows our estimations of the covariance elements for each administered dose of PMA (Eq 2.40). The covariance measurements are all positive, which is in accordance to the known biology and the positive correlation in Figure 2.13C, indeed pMEK increases the abundance ppERK. In addition, measurements of the variance indicate that ppERK is downstream of pMEK - we show that the variance of ppERK increases with covariance of pMEK and ppERK, while pMEK remains invariant. Again, this is consistent with the well established biochemistry of the MAPK pathway. Most interestingly, and similar to the *E. coli* analysis, we see that our measurements of the σ^2_{ppERK} minimally deviate from the ideal isolated two node model. Indeed, the MAPK signaling pathway is a series of activation

reactions and thus may be best represented by the cascade network. Our results show that the biochemical parameters of this pathway are set to mimic the isolated two component system [82].

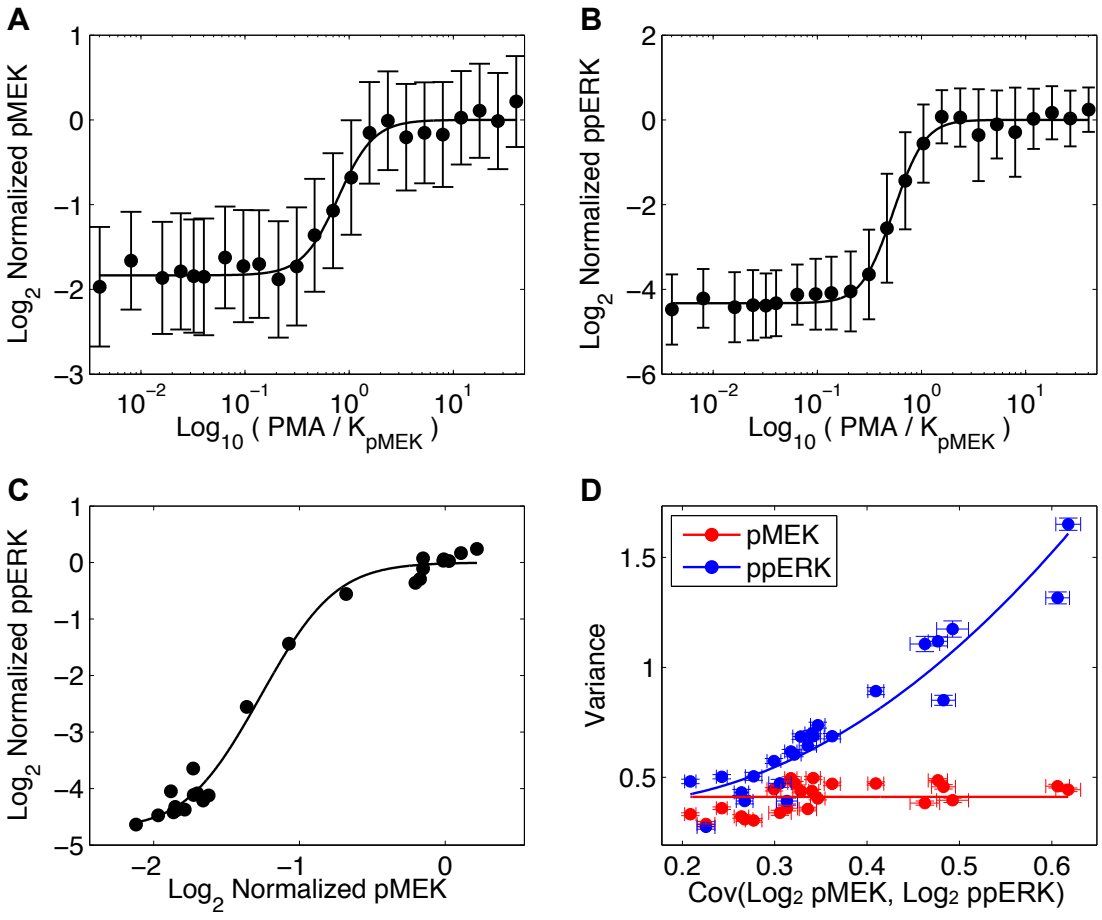

Figure 2.13: **Noise propagation in MAPK signaling in T lymphocytes.** (A) The logarithmic change in abundance of phosphorylated MEK (pS221) with increasing exogenous doses of PMA. (B) The logarithmic change in abundance of phosphorylated ERK (pT202,pY204) with increasing exogenous doses of PMA. (C) The increase of ppERK with increasing pMEK. (D) The measured $(\sigma_{\text{pMEK,ppERK}}, \sigma_i^2)$ for $i = \{\text{pMEK,ppERK}\}$ of single cell phospho-signal for various doses of PMA. Error bars are $\pm$ one standard deviation computed from 1000 bootstraps. The respective lines for pMEK and ppERK represent the best fit of our model (Eq 2.40) from the linear two component configuration.

We conclude our analysis of noise propagation in MAPK signaling of T lymphocytes by applying our directionality test. Indeed, inspection of Figure 2.14 shows

that application of the true model, pMEK → ppERK, best represents the transmission of noise. We statistically quantified this result by the previous described prescription and obtain $p = 1.5$ x 10^{-5} for $N = 24$.

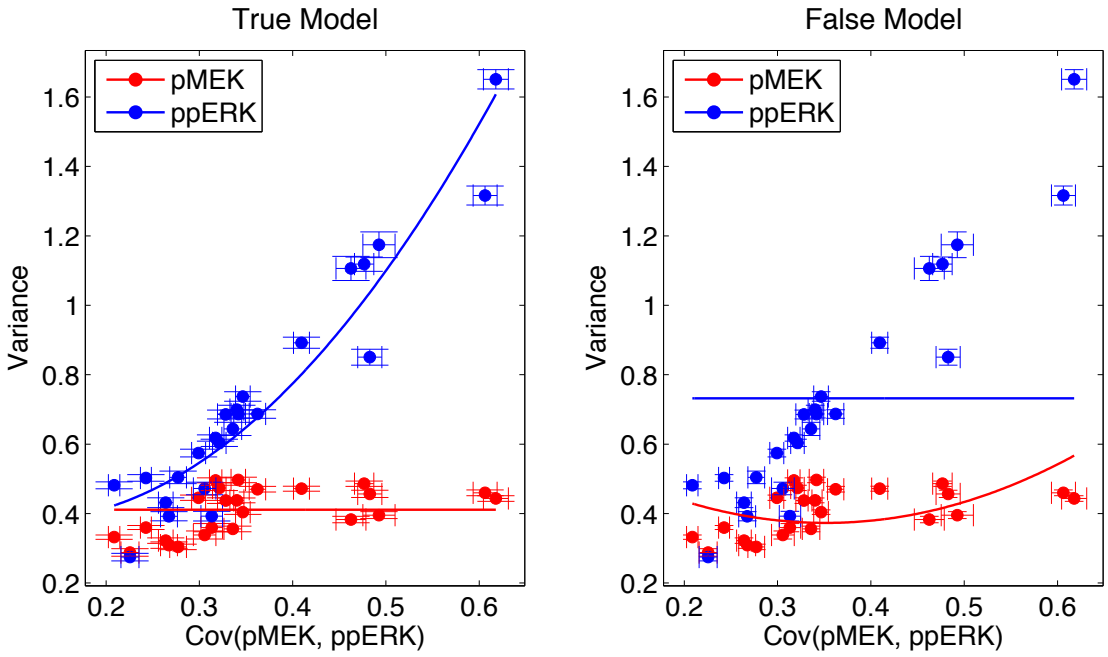

Figure 2.14: **Phospho-signaling directionality test** The covariance and variance plots with competing model fits. The true model represents a fit of covariance elements for the canonical signaling pathway, pMEK → ppERK. In contrast, the false model fits the covariance elements according to the interaction, ppERK → pMEK.

2.5.3 Constant magnitude of intrinsic noise, an experimental evaluation

We extend the results from the previous section to assess the validity of a key assumption in our model. In our derivation of the logarithmic chemical Langevin equation, we assumed that the sum of the fluctuations originating from physicochemical laws of chemical reactions and the fluctuations of parameters are con-

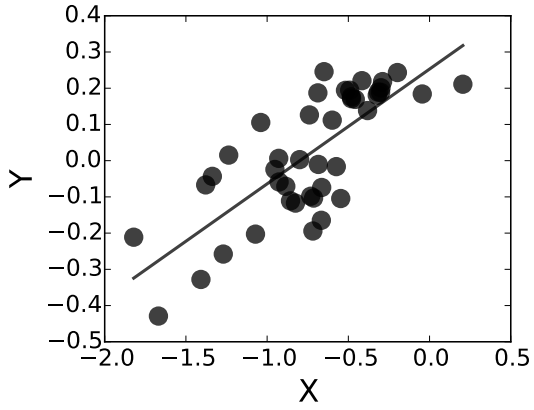

Figure 2.15: **Global noise estimate** The global noise was fit by minimizing the sum-square residuals of the data in the new coordinates (Eq 2.41). The data exhibited a correlation coefficient of 0.76.

stant (Eq 2.17). We checked this assumption by using a simple rearrangement of Eq 2.40b, resulting in

$$\underbrace{\sigma^2_{\text{ppERK},i} - 2\frac{\sigma^2_{\text{pMEK,ppERK},i}}{\sigma^2_{\text{pMEK},i}}}_{Y_i} = \beta + \eta_g \underbrace{\left(1 - 2\frac{\sigma_{\text{pMEK,ppERK},i}}{\sigma^2_{\text{pMEK},i}}\right)}_{X_i}, \tag{2.41}$$

a straight line with respect to the new coordinates (X, Y). Rewriting, we see that

$$Y_i = \beta + \eta_g X_i, \tag{2.42}$$

in which η_g represents the slope and β the Y intercept. We estimated these parameters by minimizing the sum-squared residuals (Figure 2.15). Substituting our estimated global noise, $\hat{\eta}_g$, into a rearrange expression of Equation 2.40b,

$$\beta_i = \sigma^2_{\text{ppERK},i} - \hat{\eta}_g \left(1 - 2\frac{\sigma_{\text{pMEK,ppERK},i}}{\sigma^2_{\text{pMEK},i}}\right) - 2\frac{\sigma^2_{\text{pMEK,ppERK},i}}{\sigma^2_{\text{pMEK},i}}, \tag{2.43}$$

allows us to estimate β_i for each data point. This may seem counter-intuitive, in that we just estimated $\hat{\beta}$ from our linear model (Eq 2.41). True, however, the estimated $\hat{\beta}$ from the linear fit represents an average over all samples, and our goal is to assess any systematic changes of β_i across the various stimulating conditions.

Figure 2.16 shows the resulting β_i for different ppERK activity strengths. Indeed, β_i exhibits a negligible dependency with ppERK activation strength - a linear fit of the data results in a slope of -0.03. These results suggest that our assumption of constant **Q** is representative of the intrinsic biological fluctuations of ppERK in our experimental settings.

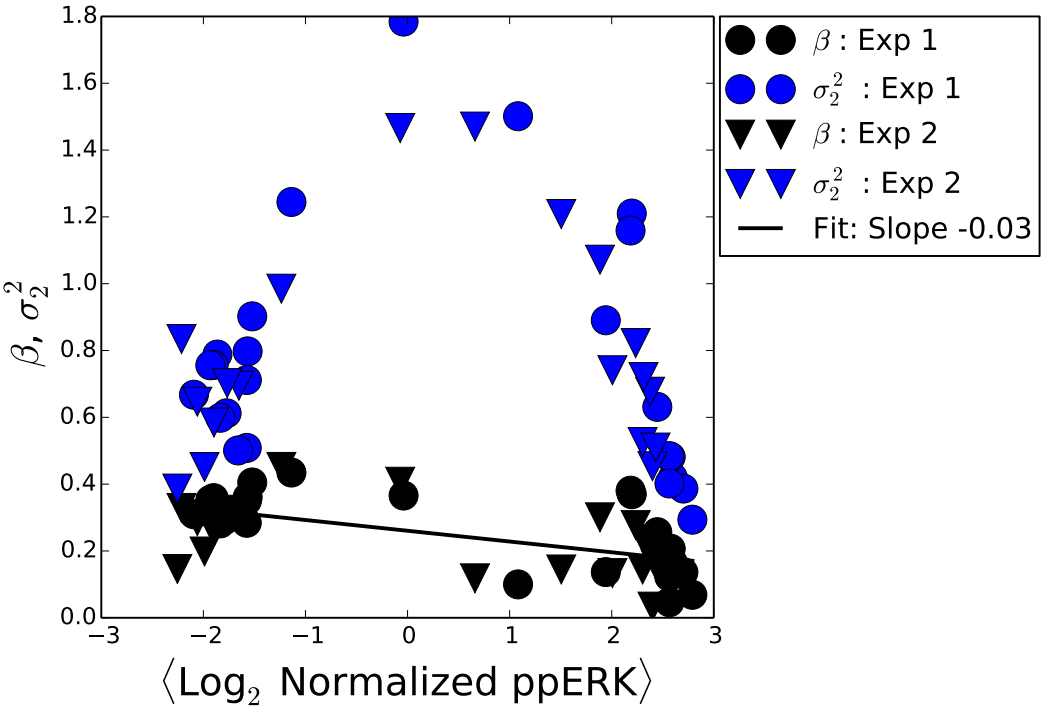

Figure 2.16: **Experimental confirmation of constant intrinsic biochemical fluctuations.** The measured variance (blue) and intrinsic biochemical fluctuations (β,black) of ppERK in T lymphocytes. Line is a linear fit of with respect to the mean activation of ppERK.

2.5.4 Identifying systems with stimulus dependent intrinsic noise

Our analysis of noise propagation is predicated on the experimentally vetted assumption that **Q** is invariant to stimulation. Indeed, such a constraint is most likely not appropriate for all biological systems. As a result we aimed at characterizing a clear diagnostic criteria to easily identify systems in which the magnitude of the noise is dependent on the stimulation strength. We limit ourselves to the isolated two component model (Eq 2.40b) and sources of noise that increase in magnitude with stimulation strength.

Let's consider this system exposed to two extreme stimulatory conditions, namely s_{min} and s_{max}, such that $\sigma_{12} \to 0$ at both locations. In Figure 2.17A we see the desired behavior, the network connectivity, and the corresponding covariance, for the extreme values of y_1 do indeed approach zero. By comparing the variances of y_2 for each condition results in

$$\sigma_2^2(s_{max}) - \sigma_2^2(s_{min}) = \beta(s_{max}) + \eta_g(s_{max}) - \beta(s_{min}) - \eta_g(s_{min}). \tag{2.44}$$

If β and η_g are constant as in our analysis, then the resulting difference is zero (Figure 2.17B). However, if β and η_g were to increase with stimulation as assumed, then the end points would not meet as depicted by Figure 2.17C. As a result, visual inspection of the covariance and variance plane is sufficient for identifying systems in which our constant noise constraint is not suitable.

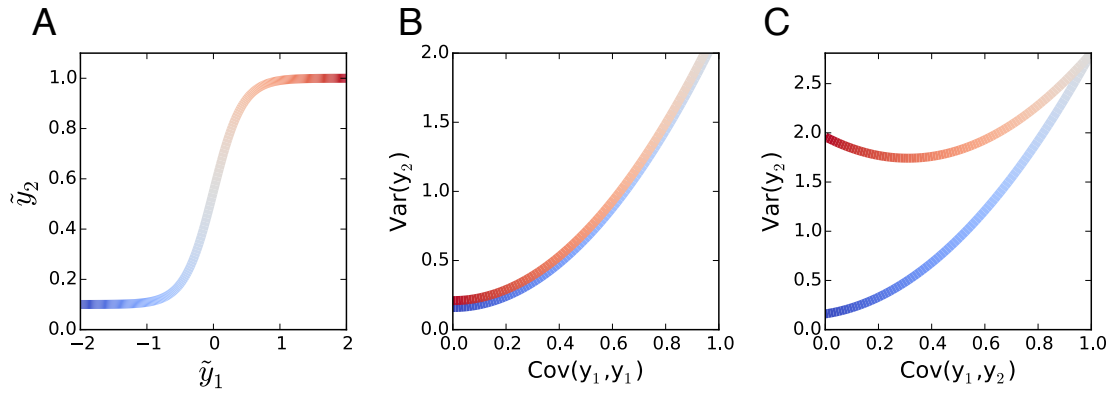

Figure 2.17: **Signatures of extrinsic noise** (A) The response of y_2 with changes of y_1 from the isolated two component model. (B) Covariance and variance plane representation of the noise propagation according to Equation 2.40b. (C) Covariance and variance plane in the isolated two node network with dose dependent fluctuations.

2.6 Summary

We experimentally uncovered that fluctuations in the abundance of biochemical species propagates according to a system of logarithmic biochemical Langevin equations. Using these equations we derived a Lyapunov equation, and as a result uncover a connection between the network connectivity $\mathbf{A}$, the estimated covariance matrix from single cell measurements $\mathbf{\Sigma}$, and the intrinsic noise matrix $\mathbf{Q}$. Using this method, we theoretically assessed the propagation of fluctuations in three biologically inspired networks: linear two node, fan-in, and a cascade. These findings were experimentally validated in a synthetic IPTG controlled *E. coli* gene regulatory network [64], and PMA activated MAPK phospho-signaling in primary mouse T lymphocytes. By leveraging the properties of noise propagation it is possible to infer the directionality of biochemical reactions from single cell data recorded at a single point in time, and measure signal originating from unmeasured signaling components. Together, we have provided a powerful tool that is built upon simple and robust calculations to study biochemical networks in living cells.

CHAPTER 3

SINGLE CELL MEASUREMENTS FOR MECHANISTIC CHARACTERIZATION OF SMALL MOLECULE CHEMICAL INHIBITORS

Individual cells rely on biochemical signaling pathways to translate environmental cues into physiological responses. Spurious activation of these pathways result in a cell's mischaracterization of environmental conditions and aberrant cellular behavior. This behavior can, in some cases, be detrimental to the health of the organism - causing ailments such as inflammatory diseases (e.g. ulcerative colitis [71]), auto-immune disorders [85, 60] and cancer [37]. A promising strategy in battling these diseases is the application of specific small molecule chemical inhibitors that target the dysregulated signaling element(s) and thereby rectify cellular behavior [56]. However, despite the tremendous success of few small-molecule therapies such as Imatinib in treating Chronic Myelogenous Leukemia [43] and Gefitnib for patients with EGFR mutant non small cell lung cancer [61, 53], the rate of success in drug development has remained incredibly slow and the field of pharmacology may need new tools to help in the development and screening of candidate compounds in more relevant settings [46, 88].

The process of drug development is organized at different levels of increasing biological scales: biomolecular, cellular, animal and lastly humans. Recent technological advances have established a pipeline to characterize *in vitro* and at the molecular level the specificity of one drug with respect to the entire human kinome [29, 45, 25]. Problems abound when attempting to expand this molecular understanding to the cellular scales. Indeed, there exist technologies such as kinobeads, that were developed to measure drug specificity and to assess response of

signaling networks to inhibition of kinases in cell lysates e.g. by quantitative mass spectrometry [8]. However, advances screening technologies at the single cell scale diverge from mechanistic understanding - focusing on automated characterization of cellular 'systems' responses [65] and behavior of phenomenologically defined populations in cell lines [75]. Yet, despite these technological advancements there still remain too many poorly performing compounds proceeding to the organismal stage of drug discovery - forcing us to re-evaluate the relevance of cell line model systems and the need for mechanistic detail in studying single cell responses (Whats wrong with drug screening today, [59]). In response, pioneering studies of *primary* cells using multidimensional single cell technologies, e.g. Flow Cytometry and Cytometry Time of Flight (CyToF) measurements, have characterized biochemical networks [40, 69] and canonical cell types response to inhibition[49, 12, 14]. While these studies have been illuminating, mechanistic principles of cellular responses to small molecule chemical inhibition has remained elusive.

Indeed the mechanistic description of response to external perturbations at the individual cell level remains an active field of investigation of practical and fundamental significance. Notably, single cell analysis of early T lymphocyte activation events showed how the biochemical organization of the T cell receptor signaling pathway can assess the quality and quantity of peptide presented [4, 32]. The mechanistic details of the proposed model led to the theoretical prediction and experimental validation of the fact that the endogenous variability of protein is sufficient to diversify individual cell responsiveness to antigen [31]. Similarly, Cotari and Voisinne *et al* [22] introduced a new methodology, Cell to Cell Variability Analysis (CCVA), to show how the diversity of IL2 Receptor α tunes the sensitivity of the individual T cell response to IL-2. A finding that has implications for the differentiation of T cells to either effector or memory phenotype during

an immune response [41]. Other examples where the study of cell-to-cell variability led to functional insight, pertain to many different biological phenomena. For example,variable abundance of the surface receptor Sca-1 in hematopoietic progenitor cells was shown to correlate with biased differentiation into erythroid or myeloid lineages [17]; diversity in the intracellular levels of phosphorylated ERK and AKT was shown to correlate with differential fates (growth versus differentiation) in PC12 cells [19], and has been implicated in the diverse time to TRAIL induced apoptosis in HELA cells [76]. For all these examples, detailed mechanistic understanding of each biological system was necessary to account for the diversity of responses to external ligands.

Measurement resolution of enzymatic states (e.g. phospho-status) at the single cell level is requisite for clear identification of properties emerging from complex nonlinear interactions between signaling network elements. Indeed, nonlinear responses often rely upon a subset of enzymes that determine the emergent behavior of the pathway. Identifying these key enzymes uncovers novel vulnerabilities of the signaling network to inhibition [11]. Examples of nonlinear responses uncovered by single cell measurements are numerous - flow cytometry measurements of ppERK accumulation in T cell receptor stimulated T lymphocytes exhibited a highly non-linear bimodal response to antigen [4]. Live single cell imaging of ppERK activation in response to growth factors was shown to be pulsatile when measured at the individual cell level [5]. Administration of either the tyrosine kinase inhibitor Gefitnib or MEK inhibitor PD325901 yielded either a frequency or a mean reduction in ppERK signaling, respectively [3]. Similarly, EGF stimulated 184A1 cells, a mammary epithelial cell line, exhibited oscillatory ERK nuclear localization (a proxy for activation) with a period invariant to EGF dose, while the number of cells exhibiting oscillatory behavior reduced markedly with increased

cell density and decreased EGF concentration [74]. These examples of the diverse complexities of biochemical signaling networks as unraveled by single-cell measurements were shown to be extremely informative when analyzing the dynamic structure of a signaling pathway. To our knowledge such combination of single cell phospho-protein measurements and mechanistic nonlinear dynamical models of signal transduction remains unexplored during drug development.

In this study, we incorporate single cell multi-parametric phospho-flow cytometry measurements, Cell to Cell Variability Analysis (CCVA, [22]), and mechanistic modeling to dissect the mechanism of action of kinase inhibitors in primary mammalian cells. We provide experimental and theoretical evidence that the network structure in which the targeted enzyme is embedded determines the signaling response to inhibition. Furthermore, we investigate the influence of protein expression variability to the sensitivity of single cells to inhibition. Lastly, we present experimental results demonstrating the functional relevance of our mechanistic models of drug inhibition on short timescales to understand drug effects on cell proliferation on longer timescales.

3.1 pSTAT5 variability in JAK inhibition response originates with the endogenous variability of STAT5

Reductionist approaches posit that properties of signal transduction pathways in living cells should be deductible from the biochemistry of its components working in concert. However, *in vitro* assays of enzyme extracts - the standard for biochemical characterization of inhibitors - do not incorporate the inherent biological complexity of cells. To illustrate this issue and establish the relevance of CCVA, we investigated the biochemistry of drug inhibition for a simple signaling pathway

- namely, JAK inhibition of the STAT5 phosphorylation response for individual T lymphocytes stimulated with the cytokine Interleukin 2 (IL-2). IL-2 stimulation induces the coalescing of the three components of the IL2 receptor [22]. Upon receptor complex formation, JAK1 and JAK3 - associated with the the intracellular domains of the receptors γ and β subunits - form an active complex (Figure 3.1A). The active complex catalyzes the phosphorylation of STAT5 (pSTAT5), which in turn initiates a large transcriptional response, including cell proliferation and inhibition of apoptosis [15, 89].

In order to monitor the inhibition of the JAK/STAT pathway at the level of individual cells, we prepared *ex vivo* mouse primary T lymphocytes and exposed them to saturating amounts of the cytokine IL-2 (2 nM) for 10 minutes followed by varied concentrations of the JAK inhibitor AZD1480 (I_{JAK}) for 15 minutes. In the absence of drug we observed that the distribution of pSTAT5 among clones is unimodal with a coefficient of variation of 77 $\pm$ 4%. The unimodal distribution pSTAT5 persisted throughout all doses of JAK inhibitor, while the geometric mean and coefficient of variation decreased monotonically with increased dosages of the inhibitory drug (Figure 3.1B). We then measured and tested the average response of cells to JAK inhibition with an inhibitory Hill function, we measured a half effective inhibition concentration (IC_{50}) of 8.2 $\pm$ 0.5 nM and set the Hill coefficient to $n_{\text{Hill}} = 1$. This lack of cooperativity strongly suggests that JAK inhibition is a single step process, despite the intrinsic biological complexity in JAK-STAT activation . In addition, the monotonic decrease in the CV_{pSTAT5} suggested that the variation in pSTAT5 among single cell clones does not originate from stochasticity of chemical reactions with small numbers of reactants. If this were the mechanism of our observed diversity, the variance of pSTAT5 abundance per cell would scale as the inverse square root of the mean pSTAT5 abundance [84, 35]. This would result

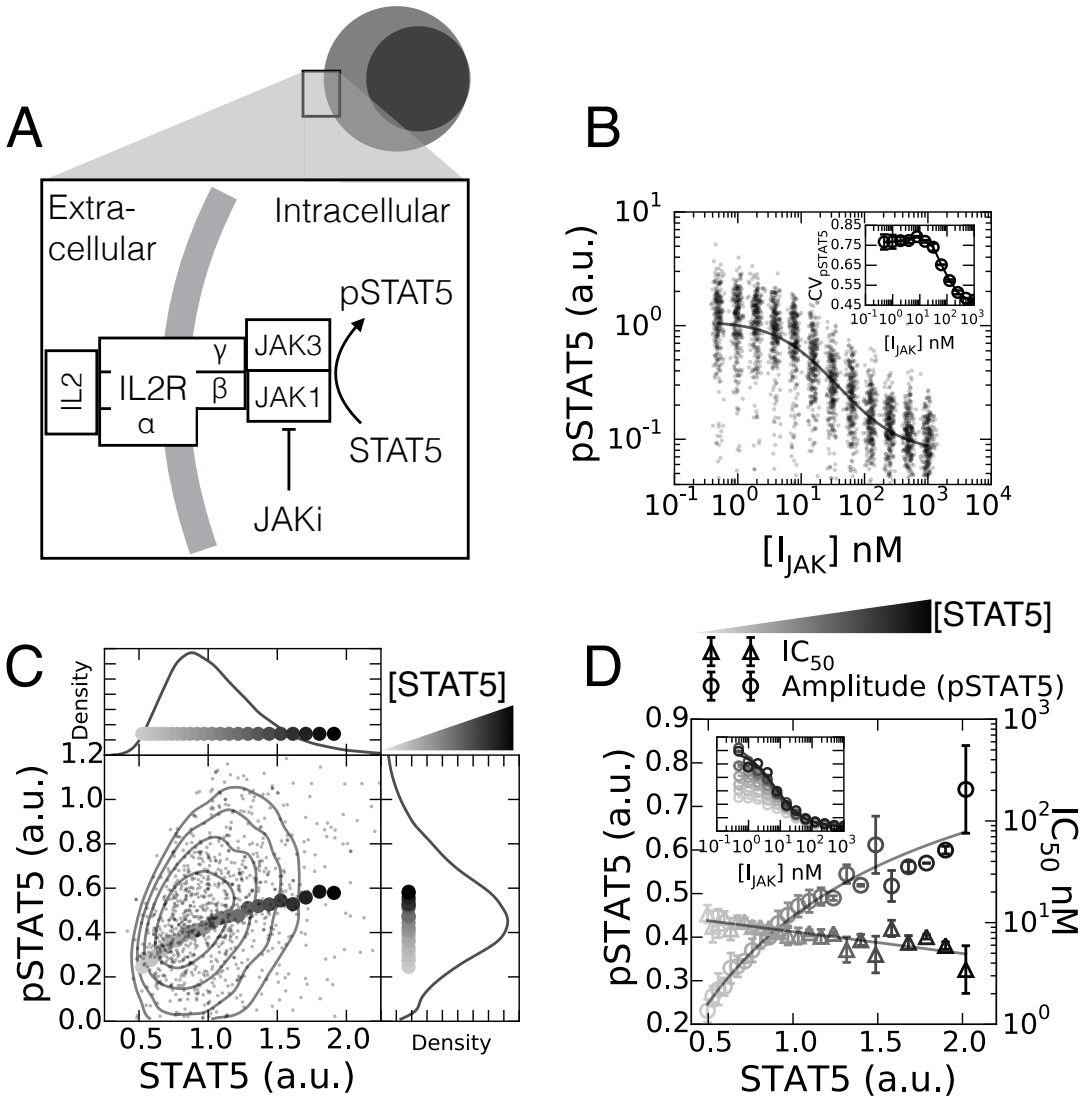

Figure 3.1: **Variability of endogenous protein abundance correlates with single cell response to chemical inhibition.** (A) IL-2 stimulation of the JAK-STAT pathway. (B) Single cell pSTAT5 abundance in response to JAK inhibitor AZD1480. Inset, the coefficient of variation (CV) response to inhibition. (C) Single cell distribution of the total abundance of STAT5 and pSTAT5 in cells not treated with inhibitor, $[\mathrm{I_{JAK}}] = 0$. Curve shows the resulting geometric mean of the pSTAT5 abundance conditioned on STAT5 abundance per cell. (D) Cell-to-Cell Variability Analysis reveals that the pSTAT5 response amplitude is correlated with STAT5 abundance and the sensitivity of cells to inhibition is slightly negatively correlated (error bars are standard deviation of experimental duplicates).

in an increase CV_{pSTAT5} with increase JAK inhibitor concentration. Together these data suggest a simple biochemical mechanism of inhibition of single cell signaling and that individual clones have unique internal parameters that manifest as diverse levels of pSTAT5.

We then focused on elucidating a source of pSTAT5 variability in response to JAK inhibition. Having ruled out the stochasticity introduced from the physical properties of biochemical reactions, we decided to measure an observable that would complement the CCVA framework. We chose to monitor the abundances of both total STAT5 and pSTAT5 per cell by phospho-flow cytometry [31, 22]. Using our single cell measurements and CCVA we performed *in silico* STAT5 titration experiments, and found that the geometric mean of pSTAT5 correlates with STAT5 abundance in the absence of JAK inhibitor (Figure 3.1C). Extending this finding, we investigated both the JAK inhibitor dose response amplitude and half effective inhibitor concentration (IC_{50}) with varying abundance of STAT5. Indeed, the amplitude of pSTAT5 response increased with STAT5 expression as a Hill function with a Hill coefficient of 1, while the IC_{50} reduced exponentially with a scale of approximately -2.02 (STAT5 a.u., Figure 3.1D). Monitoring the extent of drug inhibition at the single cell level enabled us to establish new experimental observations regarding signal inhibition *e.g.* by correlating the cell-to-cell variability of the response with the heterogeneity of expression of signaling components.

Our preliminary analysis provided detailed quantitative observation regarding the dependence of pSTAT5 to the endogenous variability of STAT5 abundance, however it lacks a simple quantitative description of the biochemistry of inhibition in live cells. To do this we tested three simple biochemical models that may account for the transmission of STAT5 variability to pSTAT5 levels per cell, and the

biochemical mechanism of JAK inhibition by the AZD1480 inhibitor - namely noncompetitive, uncompetitive, and competitive inhibition [52, 20]. In the proceeding sections we describe, in detail, our derivation for equations of each inhibitor model and our model selection procedure.

3.1.1 Noncompetitive inhibition

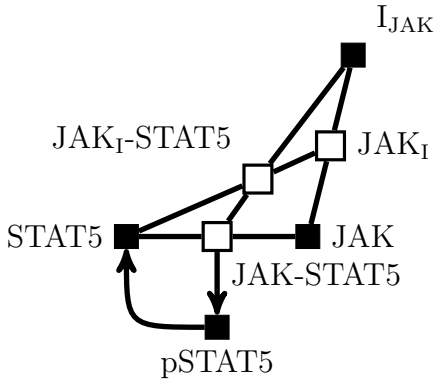

Figure 3.2: **Noncompetitive inhibitor model** Graphical representation of the noncompetitive JAK inhibitor model. The square symbols represent either a chemical species (filled) or a chemical complex (open). The interaction between chemical species and complexes is either an equilibrium reaction (lines) or unidirectional catalyzed reaction (arrows).

Noncompetitive inhibitors bind to their target enzyme irrespective to the presence of the enzymes substrate. Here, we consider the enzyme substrate to be the measurable abundance of the STAT5 protein. Although, ATP levels are an important consideration, measurements of single cell ATP are beyond the schope of this work.

To develop a quantitative description of noncompetitive inhibition in living cells we start by considering the set of mass action chemical kinetic equations governing

the system:

$$\frac{d}{dt}[\text{pSTAT5}] = \Gamma[\text{JAK-STAT5}] - \lambda[\text{pSTAT5}], \tag{3.1a}$$

$$\frac{d}{dt}[\text{JAK}_\text{I}\text{-STAT5}] = k_+^{\text{I}_\text{JAK}}[\text{I}_\text{JAK}][\text{JAK-STAT5}] - k_-^{\text{I}_\text{JAK}}[\text{JAK}_\text{I}\text{-STAT5}] + k_+[\text{JAK}_\text{I}][\text{STAT5}] - k_-[\text{JAK}_\text{I}\text{-STAT5}], \tag{3.1b}$$

$$\frac{d}{dt}[\text{JAK-STAT5}] = k_+[\text{JAK}][\text{STAT5}] - k_-[\text{JAK-STAT5}] - \Gamma[\text{JAK-STAT5}] + k_-^{\text{I}_\text{JAK}}[\text{JAK}_\text{I}\text{-STAT5}] - k_+^{\text{I}_\text{JAK}}[\text{I}_\text{JAK}][\text{JAK-STAT5}]. \tag{3.1c}$$

Where we identify the equilibrium reaction parameters as denoted as k_+ for the association of JAK and STAT5, k_- represents the dissociation of the JAK-STAT5 complex, and $k_\pm^{\text{I}_\text{JAK}}$ representing the inhibitor forward and reverse rate constants. The enzymatic reactions of the phosphorylation and dephosphorylation are parameterized by rate constants Γ and λ, respectively. Conservation of mass dictates that,

$$[\text{STAT5}_\text{total}] = [\text{pSTAT5}] + [\text{STAT5}] + [\text{JAK-STAT5}] + [\text{JAK}_\text{I}\text{-STAT5}], \tag{3.2a}$$

$$[\text{JAK}_\text{total}] = [\text{JAK}] + [\text{JAK}_\text{I}] + [\text{JAK-STAT5}] + [\text{JAK}_\text{I}\text{-STAT5}]. \tag{3.2b}$$

However, we simplify the STAT5 constraint by considering the relative abundance of both JAK and STAT5 in our experimental system. In our system, Feinerman *et al* [30] previously presented values of the abundance of $[\text{STAT5}_\text{total}]$ that are nearly an order of magnitude greater than the IL2 Receptor β chain. The IL2Rβ chain is important for it is the limiting component of the active IL2 Receptor, JAK1, and JAK3 complex [30]. As a result, the total effective abundance of active JAK in our system ($[\text{JAK}_\text{total}]$) is negligible when considered with the STAT5 abundance - motivating our approximation of $[\text{STAT5}_\text{total}] \approx [\text{pSTAT5}] + [\text{STAT5}]$.

In addition we simplify the dynamic equations by using a quasi-steady state approximation, which assumes that the binding / debinding kinetics of the equi-

librium reactions are much faster than the catalytic reactions. This separation of time-scales between the equilibrium reactions and the catalytic reactions simplifies the dynamics by setting all the time derivatives of the equilibrium reactions to zero. Using the quasi-steady state approximation in conjunction with our approximated constraints we obtain,

$$\frac{1}{\Gamma}\frac{d}{dt}[\text{pSTAT5}] = \frac{[\text{JAK}_{\text{total}}][\text{STAT5}]}{[\text{STAT5}](1+\mathcal{I}_J) + k_m(1+\mathcal{I}_J\frac{k_d}{k_m})} - \frac{\lambda}{\Gamma}[\text{pSTAT5}]. \tag{3.3}$$

Here $\mathcal{I}_J = \frac{k_+^{\text{IJAK}}}{k_-^{\text{IJAK}}}[\text{I}_{\text{JAK}}]$, $k_d = \frac{k_-}{k_+}$, and lastly $k_m = \frac{\Gamma+k_-}{k_+}$. Equation 3.3 can be simplified further by considering $\frac{k_d}{k_m}$ in terms of the rate constants

$$\frac{k_d}{k_m} = \frac{1}{1+\underbrace{\frac{\Gamma}{k_-}}_{\approx 0}} \approx 1. \tag{3.4}$$

Incorporating this approximation and noting that Γ sets the time-scale, we change to dimensionless units, $d\tau = \Gamma dt$ and $\Lambda = \frac{\lambda}{\Gamma}$, so that

$$\frac{d}{d\tau}[\text{pSTAT5}] = \frac{[\text{JAK}_{\text{total}}]}{(1+\mathcal{I}_J)}\frac{[\text{STAT5}]}{[\text{STAT5}]+k_m} - \Lambda[\text{pSTAT5}]. \tag{3.5}$$

Our experimental measurements are at steady state for pSTAT5, and are obtained as arbitrary fluorescent units instead of concentrations as required by Equation 3.5. Therefore, setting the derivative to zero, normalizing by k_m, and introducing the parameter ϕ we arrive at our dimensionless model equation

$$y = \frac{\alpha}{(1+\mathcal{I}_J)}\frac{\phi x - y}{(\phi x - y)+1}. \tag{3.6}$$

Where $y = [\text{pSTAT5}]/k_m$ represents our pSTAT5 measurement, $\phi x = [\text{STAT5}_{\text{total}}]/k_m$ is the adjusted total STAT5 measurement, and $\alpha = \frac{[\text{JAK}_{\text{total}}]}{\Lambda k_m}$ representing the our pSTAT5 amplitude measurements. The parameter ϕ is necessary because we apply the constraint of mass conservation to our dynamic equations, without measuring our observables in units of absolute number of molecules.

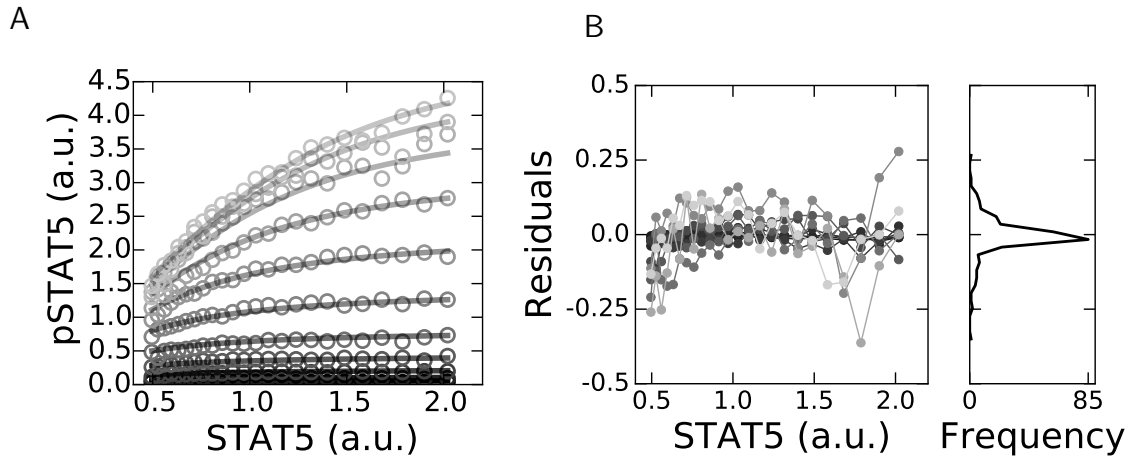

Figure 3.3: **Noncompetitive inhibitor fit to data.** (A) The resulting fit of the data (open circles) with the noncompetitive model (lines) presented in Equation 3.7. (B) The residuals of the data points from the noncompetitive model fit for each dose of JAK inhibitor. The color gradient from gray to black represents increasing AZD1480 concentration, I_{JAK}.

To test the mechanism of inhibition we fit our data to the closed form solution

$$\boxed{y = \frac{1}{2}\left[\phi x + 1 + \frac{\alpha}{1+\mathcal{I}_J} - \sqrt{\left(\phi x + 1 + \frac{\alpha}{1+\mathcal{I}_J}\right)^2 - 4\alpha\frac{\phi x}{1+\mathcal{I}_J}}\right]}. \tag{3.7}$$

Figure 3.3 demonstrates the excellent agreement of our model with the experimental data.

3.1.2 Uncompetitive inhibition

The uncompetitive inhibitor is a small molecule that binds exclusively to its target enzyme while bound to its cognate substrate - in this case the JAK-STAT5 chemical complex (Figure 3.4). Indeed, this mechanism of inhibition manifests in the equilibrium equations of the drug-enzyme interaction and the conservation equation for $[JAK_{total}]$. Apart from the inhibitor's dynamic equations our analysis

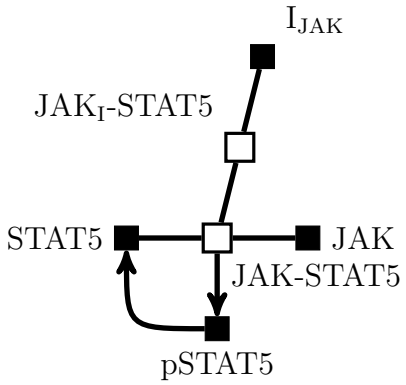

Figure 3.4: **Uncompetitive inhibitor model.** Graphical representation of the uncompetitive JAK inhibition model.

remains the same as the noncompetitive inhibitor scenario,

$$\frac{d}{dt}[\text{pSTAT5}] = \Gamma[\text{JAK-STAT5}] - \lambda[\text{pSTAT5}], \tag{3.8a}$$

$$\frac{d}{dt}[\text{JAK-STAT5}] = k_+[\text{STAT5}][\text{JAK}] - k_-[\text{JAK-STAT5}] - \Gamma[\text{JAK-STAT5}] + k_-^{\text{I}_{\text{JAK}}}[\text{JAK}_\text{I}\text{-STAT5}] - k_+^{\text{I}_{\text{JAK}}}[\text{I}_{\text{JAK}}][\text{JAK-STAT5}], \tag{3.8b}$$

$$\frac{d}{dt}[\text{JAK}_\text{I}\text{-STAT5}] = k_+^{\text{I}_{\text{JAK}}}[\text{I}_{\text{JAK}}][\text{JAK-STAT5}] - k_-^{\text{I}_{\text{JAK}}}[\text{JAK}_\text{I}\text{-STAT5}]. \tag{3.8c}$$

The following mass conservation constraint follows as:

$$[\text{STAT5}_{\text{total}}] \approx [\text{pSTAT5}] + [\text{STAT5}], \tag{3.9a}$$

$$[\text{JAK}_{\text{total}}] = [\text{JAK}] + [\text{JAK-STAT5}] + [\text{JAK}_\text{I}\text{-STAT5}]. \tag{3.9b}$$

Then applying our quasi-steady state approximation as before we obtain an expression for the dynamics of pSTAT5,

$$\frac{d}{d\tau}[\text{pSTAT5}] = [\text{JAK}_{\text{total}}]\frac{[\text{STAT5}]}{[\text{STAT5}](1+\mathcal{I}_J) + k_m} - \Lambda[\text{pSTAT5}]. \tag{3.10}$$

Where $\mathcal{I}_J = \frac{k_+^{\text{I}_{\text{JAK}}}[\text{I}_{\text{JAK}}]}{k_-^{\text{I}_{\text{JAK}}}}$, $k_m = \frac{k_- + \Gamma}{k_+}$, and $[\text{STAT5}] = [\text{STAT5}_{\text{total}}] - [\text{pSTAT5}]$. Solving for the unitless variables describe in Equation 3.6, we obtain

$$y = \frac{\alpha}{1+\mathcal{I}_J}\ \frac{\phi x - y}{\phi x - y + \frac{1}{1+\mathcal{I}_J}}. \tag{3.11}$$

This representation provides a transparent understanding of the difference between the noncompetitive and uncompetitive inhibitor. Unlike the noncompetitive inhibitor which reduces the response amplitude exclusively (*i.e.* α), the uncompetitive inhibitor effectively reduces both the response amplitude and the half effective inhibition concentration.

We conclude by presenting the closed form solution used to fit the data (Figure 3.5)

$$\boxed{y = \frac{1}{2}\left[\phi x + \frac{1+\alpha}{1+\mathcal{I}_J} - \sqrt{\left(\phi x + \frac{1+\alpha}{1+\mathcal{I}_J}\right)^2 - 4\alpha\frac{\phi x}{1+\mathcal{I}_J}}\right]}. \tag{3.12}$$

In Figure 3.5 we see that the model introduces systematic error in estimating pSTAT5 levels for cells with low expression of STAT5.

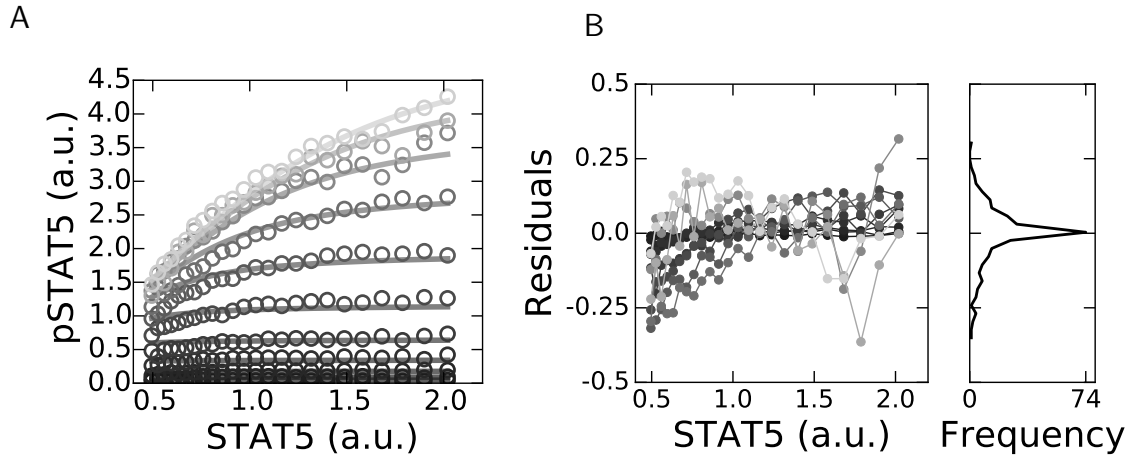

Figure 3.5: **Uncompetitive inhibitor model fit.** (A) The resulting fit of the data (open circles) and the uncompetitive model (lines) represented by Equation 3.12. (B) The residuals of the data points from the uncompetitive model fit for each dose of JAK inhibitor. The color gradient from gray to black represents increasing AZD1480 concentration, I_{JAK}.

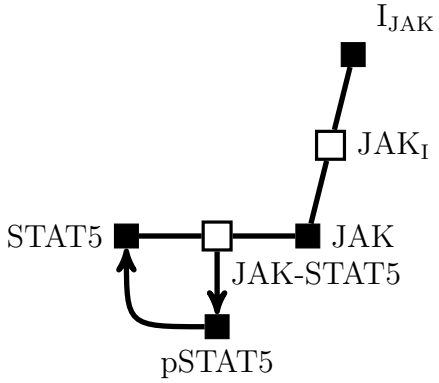

Figure 3.6: **Competitive inhibitor model.** Graphical representation of the competitive inhibition model.

3.1.3 Competitive inhibition

Competitive inhibitors compete with the substrate for binding to the target enzyme. Apart from the difference in the dynamic equations describing the unique mechanism of action, the proceeding analysis utilizes the same approximations as the previous two sections. Incorporating the inhibitor mechanism of action into the dynamics

$$\frac{d}{dt}[\text{pSTAT5}] = \Gamma[\text{JAK-STAT5}] - \lambda[\text{pSTAT5}], \tag{3.13a}$$

$$\frac{d}{dt}[\text{JAK-STAT5}] = k_+[\text{STAT5}][\text{JAK}] - k_-[\text{JAK-STAT5}] - \Gamma[\text{JAK-STAT5}], \tag{3.13b}$$

$$\frac{d}{dt}[\text{JAK}_\text{I}\text{-STAT5}] = k_-^{\text{I}_{\text{JAK}}}[\text{JAK}_\text{I}] - k_+^{\text{I}_{\text{JAK}}}[\text{I}_{\text{JAK}}][\text{JAK}]. \tag{3.13c}$$

As before we approximate the mass conservation equations of constraint as

$$[\text{STAT5}_{\text{total}}] = [\text{pSTAT5}] + [\text{STAT5}], \tag{3.14a}$$

$$[\text{JAK}_{\text{total}}] = [\text{JAK}] + [\text{JAK-STAT5}] + [\text{JAK}_\text{I}]. \tag{3.14b}$$

Applying our quasi-steady state approximation

$$\frac{d}{d\tau}[\text{pSTAT5}] = [\text{JAK}_{\text{total}}]\frac{[\text{STAT5}]}{[\text{STAT5}] + k_m(1+\mathcal{I}_J)} - \Lambda[\text{pSTAT5}], \tag{3.15}$$

with $\mathcal{I}_J = \frac{k_+^{\mathrm{I_{JAK}}}[\mathrm{I_{JAK}}]}{k_-^{\mathrm{I_{JAK}}}}$, $k_m = \frac{k_- + \Gamma}{k_+}$, and $[\mathrm{STAT5}] = [\mathrm{STAT5_{total}}] - [\mathrm{pSTAT5}]$. As before we proceed by solving for the non-dimensional variables describe in Equation 3.6 and arrive at

$$y = \alpha \frac{\phi x - y}{(\phi x - y) + 1 + \mathcal{I}_J}. \tag{3.16}$$

From this equation we can clearly identify the unique functional difference between the noncompetitive and uncompetitive inhibitors with the competitive inhibitor. Unlike the former two equations (Eqs 3.6,3.11) representing the JAK inhibitor action, the competitive inhibitor increases the half effective concentration of STAT5.

The closed form solution used to fit our data is

$$\boxed{y = \frac{1}{2}\left[\phi x + \alpha + 1 + \mathcal{I}_J - \sqrt{(\phi x + \alpha + 1 + \mathcal{I}_J)^2 - 4\alpha\phi x}\right]}. \tag{3.17}$$

Figure 3.7, shows the large systematic error of this model in approximating our experimental data.

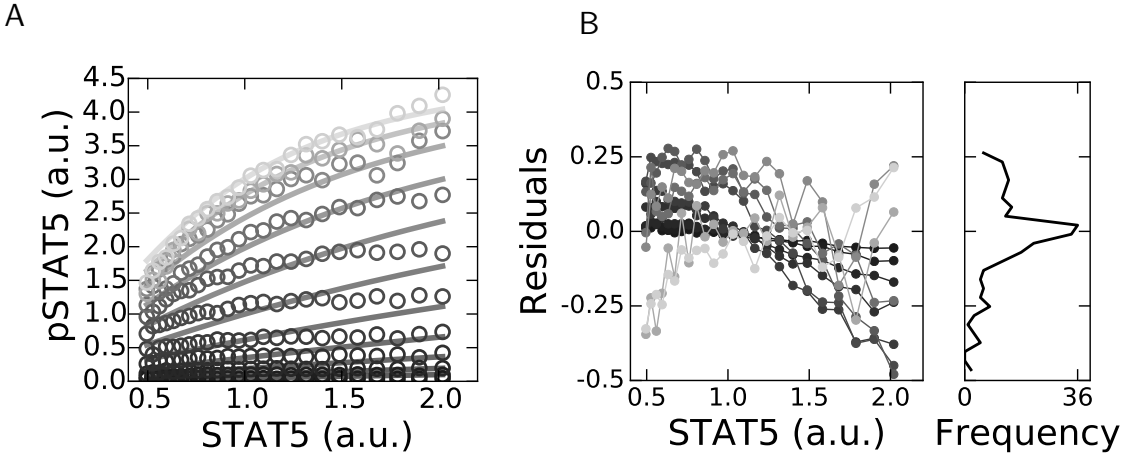

Figure 3.7: **Competitive inhibitor model fit.** (A) The resulting fit of data (open circles) and the competitive inhibitor model model represented in Equation 3.17 (lines). (B) The residuals of the data points from the uncompetitive model fit for each dose of JAK inhibitor. The color gradient from gray to black represents increasing AZD concentration, $\mathrm{I_{JAK}}$.

3.1.4 Model selection

To find the most likely mechanism of action of AZD1480 in living cells we measured the ability of each possible model to explain our data. Each model is represented as a function f_i, for $i = \{\text{Noncompetitive, Uncompetitive, Competitive}\}$, derived for each mechanism of inhibition. We measured the sum of squared residuals (ϵ_i) for each model with respect to our data,

$$\epsilon_i = \sum_{l=1}^{N}\sum_{m=1}^{M}\left(y_{l,m} - f_i\left([\mathrm{I_{JAK}}]_l, x_m; \vec{\theta}_i\right)\right)^2 \qquad \text{where } \vec{\theta}_i = \{\hat{\alpha}_i, \hat{\phi}_i, \hat{k_{I_i}}\}, \tag{3.18}$$

to obtain an estimate of the goodness of fit. In this equation N is the number of inhibitor doses and M is the number of CCVA determined STAT5 data points. The mechanism of inhibition is determined by selecting the model function, and the corresponding optimal parameters $\vec{\theta}_i$, that exhibit the smallest sum of squared residuals.

There exists several methods for optimal model selection; we chose the minimization of the sum of squared residuals as our criteria because of the simplicity of our considered models. Specifically, each model that was considered consisted of the same number of parameters (complexity) making techniques aimed at quantitatively compromising between model complexity with statistical accuracy not necessary (*e.g.* BIC and AIC).

We found that the noncompetitive inhibition model (Equation 3.7) for AZD1480 action best described our experimental observations (smallest ϵ, Figure 3.8A). Biologically, this mechanism is intuitive, because this inhibitor has been documented to compete with ATP for occupancy of the ATP binding pocket of JAK and not STAT5 [55]. We further quantified the agreement of our model fit by measuring a coefficient of determination (R^2) of 0.95 (Figure 3.8B) of data linearized according

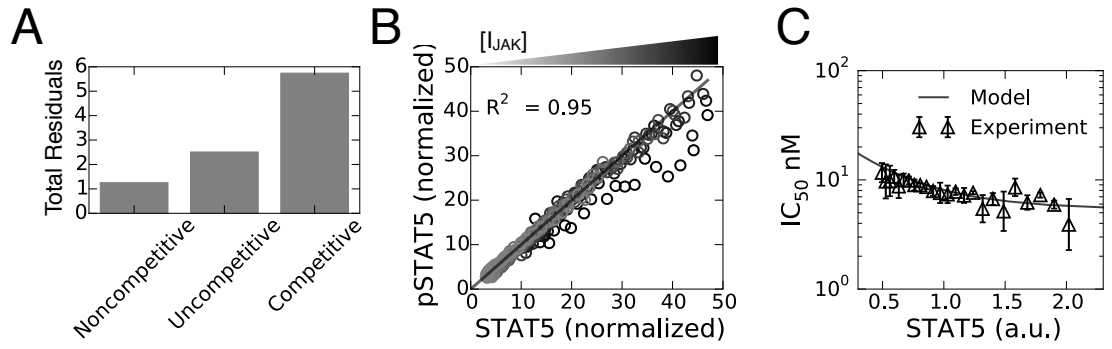

Figure 3.8: **CCVA reveals the most likely mechanism of AZD1480 in live single cells.** (A) Each model was tested against our data by measuring the sum of squared residuals of our model predictions to the data points (total residuals). (B) Linearized data according to the coordinate transform (Eq 3.19) of the optimal model and estimated parameters reveal excellent agreement between model (line) and data (open circles), $R^2 = 0.95$. (C) Overlay of measured IC_{50} with respect to STAT5 abundance as measured from CCVA analysis of data (triangles, error bars standard deviation experimental duplicates) and predicted by our optimal model (line).

to

$$\underbrace{y_{\text{data}}\left(1+\hat{\mathcal{I}}_J\right)\left(\hat{\phi}x_{\text{data}}-y_{\text{data}}+1\right)}_{\text{Normalized pSTAT5}}=\underbrace{\hat{\alpha}(\hat{\phi}x_{\text{data}}-y_{\text{data}})}_{\text{Normalized STAT5}}. \tag{3.19}$$

Eq 3.19 represents a rearrangement of Eq 3.6 for parameter values estimated from the data (Table 3.1). Lastly, we validated that our model could account for the small dependence of the IC_{50} on STAT5 expression. We see excellent agreement between our IC_{50} measurements in Figure 3.1D with the estimated IC_{50} from our model (Figure 3.8C).

Table 3.1: **Noncompetitive inhibitor parameter estimates** Parameters obtained from least-square fits of Equation 3.7 to our experimental data.

Parameter	Interpretation	Rep 1	Rep 2	Mean
ϕ	prefactor for total STAT5 expression	4.13	5.05	4.59
α	normalized amplitude	5.65	5.32	5.49
k_I	inhibitor dissociation constant	5.33 nM	4.93 nM	5.13 nM

3.2 Inhibition of TCR induced MAPK in T lymphocytes

We investigated the inhibition of a large signaling cascade, namely antigen-driven MAP kinase activation in primary mouse T lymphocytes (T cells). Upon exposure with activating ligands (*e.g.* complex of a peptide with Multi-Histocompatibility Complex on the surface of antigen presenting cells), T cells activate their receptors (TCR), through activation of a SRC Family Tyrosine Kinase (SRC) (*e.g.* Lck a SRC family kinase bound to CD8), which, in turns, trigger a cascade of kinase activation leading to ERK phosphorylation (Figure 3.9). We chose this model system because its complex network topology and its functional relevance: aberrant activation in the ERK pathway is often involved with oncogenesis [37], making it a key pathway to be targeted with drug inhibitors in multiple tumor settings.

3.2.1 Single cell measurements reveal diverse modes of inhibition

We decompose TCR mediated activation of ERK into two smaller subnetworks: 1) a receptor proximal signaling cascade with positive and negative feedback regulation, and 2) the unidirectional MAP kinase (MAPK). We hypothesized that applying inhibitors that target enzymes specific to each signaling subnetwork will produce unique responses in terms of ERK phosphorylation.

We tested our hypothesis by administering inhibitors targeting the two signaling subnetworks - a SRC inhibitor (Dasatinib) for the receptor proximal component and a MEK inhibitor (PD325901) for the MAPK component - in activated T cells. Measuring the average response of cells to both inhibitors (Figure 3.10A) results

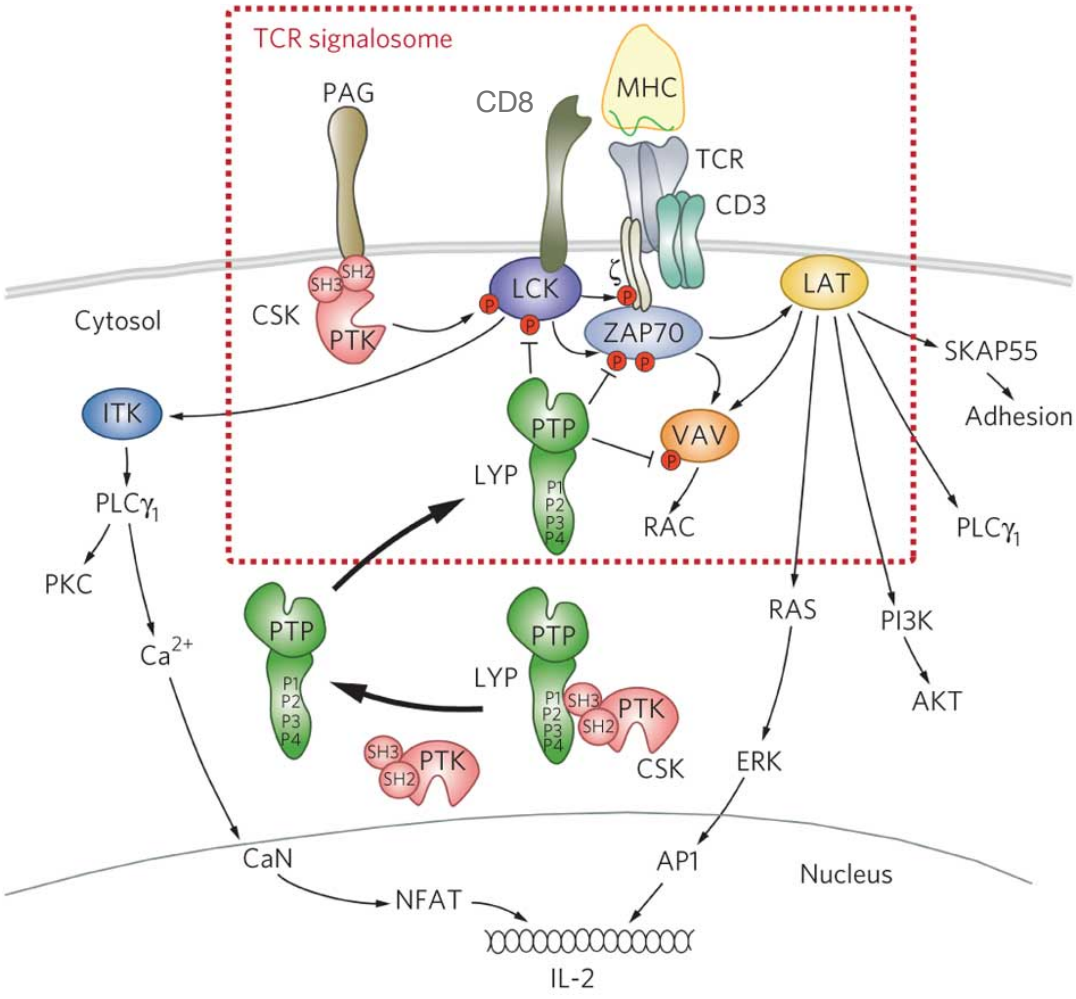

Figure 3.9: **TCR signaling biology.** Reprinted by permission from Macmillan Publishers Ltd: Nature Chemical Biology [86], copyright (2012).

in amplitude reduction and a trivial inhibition model. However, upon applying our single cell methodology (as introduced in the previous section), we found that ppERK signaling exhibits an all or none (bimodal) response (Figure 3.10B)- consistent with previous studies of T lymphocyte [4] and MCF10A cell activation [3]. Analyzing the single-cell response to Dasatinib, the SRC inhibitor, shows a bimodal response - individual cells transition from maximal to undetectable quantities of ppERK (Figure 3.10B,D). Meanwhile, application of PD325901, the MEK inhibitor, exhibits an analog response - a reduction in ppERK amplitude with dose (Figure 3.10C,E). Dasatinib and PD325901 exhibit differential modes of ppERK inhibition.

Measurements of the drug response by canonical parameters - IC_{50} and amplitude - at the bulk level failed to reveal disparate modes of inhibition as observed at the individual cell level (Figures 3.10D,E). We characterized the two modes of inhibition by fitting the distribution of ppERK abundance per cell to a Gaussian mixture model, whose statistics can be summarized with two parameters - μ_+ for the mean logarithmic measurement of ppERK among activated cells and α_+ for the fraction of activated cells (Figure 3.10F) - at each dose of inhibitor. In Figure 3.10G, we report that MEK inhibitor operates solely upon the mode of ppERK abundance among activated cells, μ_+ and acts as a analog inhibitor of ERK activation; SRC inhibitor upon the fraction of active cells, α_+ and acts as a digital inhibitor. Similar results were obtained using genetically-modified cell lines that express reporters of ERK activity; our single-cell phospho-profiling methodology has the added advantage of being applicable to primary cells (*e.g.* taken from patients), without added modification. Hence, by utilizing single-cell measurements, we were able to demonstrate that there exist two modes of inhibition in the MAPK signaling cascade: analog (when μ_+ varies with drug concentration)

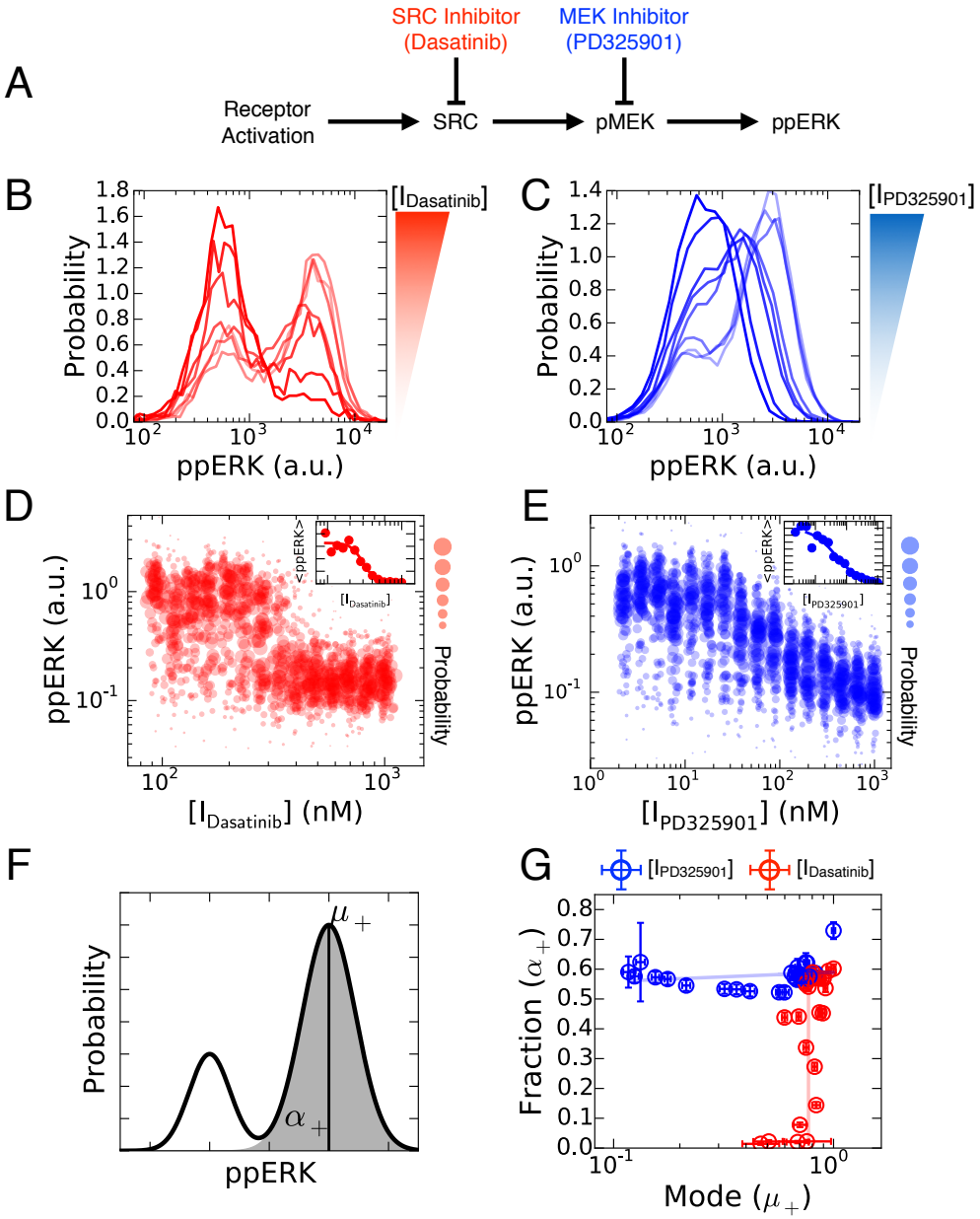

Figure 3.10: **Inhibitor target specific modes of inhibition.** (A) The implicit model of inhibitor action. (B) Histograms of single T cell response to simultaneous TCR stimulation and SRC inhibition with Dasatinib. (C) Histograms of single T cell response to simultaneous TCR stimulation and MEK inhibition with PD325901. (D) Single cell Dasatinib dose response of TCR activated T cells - inset represents the mean response of all cells. (E) Single cell PD325901 dose response of TCR activated T cells - inset represents the mean response of all cells. (F) Single cell data were modeled as a mixture of Gaussian distributions, where α_+ represents the fraction of activated cells, and μ_+ represents the average abundance of ppERK among activated cells. (G) The (μ_+,α_+) plane shows the orthogonal modes of inhibition (error bars are standard error of mean from 100 samples of 500 T cells per dose of inhibitor chosen randomly and with replacement).

or digital (when it is α_+ that varies with drug concentration). Both modes of inhibition can be associated with different enzymes at different stages of the ERK cascade.

3.2.2 Coarse grained model of TCR activation of ERK

We next sought to test whether the enzymatic context, *i.e.* the position of the enzyme undergoing inhibition within the transduction cascade, would explain the two distinct modes of inhibition (analog vs digital) observed in Figure 3.10. We developed a coarse grained model accounting for ERK phosphorylation downstream of SRC activation [4, 31, 32] in order to generate mechanistic understanding and actionable predictions to be validated. Our model explicitly incorporates measurable quantities, control parameters, and the inhibitor targeted species while coarsening the large biological complexity of this signaling cascade into a phenomenological biochemical species of activated SRC associated with the engaged ligand/receptor complex (denoted SRC*). The graphical representation of our model (Figure 3.11) emphasizes the two subnetworks at stake here : 1) active SRC (SRC*) is controlled by competing positive and negative feedbacks, and 2) MEK and ERK are activated upon formation of SRC* in a unidirectional manner (no feedback). These modeling components encompass key features of ERK activation in the context of antigen activation in T lymphocytes [4, 31, 32], without loss of generality.

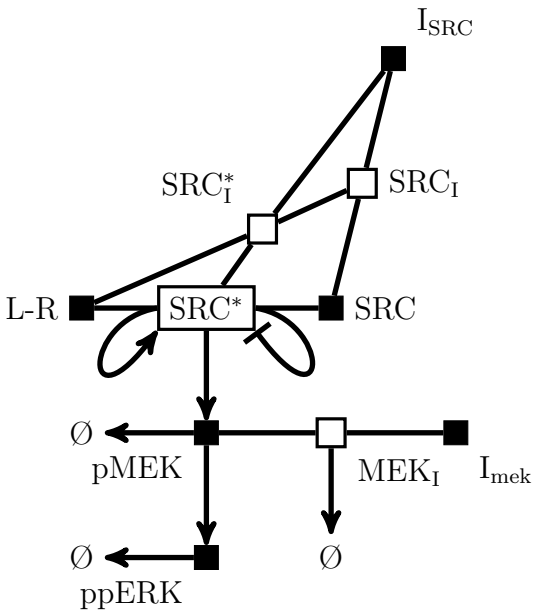

Figure 3.11: **Coarse-grained TCR signaling model.** Empty squares designate chemical complexes, while filled squares are independent chemical species. Arrow-less lines connecting two squares indicates an equilibrium reaction, while lines with arrows indicate enzymatic reactions.

The components interact according to mass action differential equations:

$$\frac{d}{dt}[\mathrm{SRC}^*] = k_+[L-R][\mathrm{SRC}] - k_-[\mathrm{SRC}^*] + k_+[L-R][\mathrm{SRC}]\frac{\Gamma_+[\mathrm{SRC}^*]^n}{[\mathrm{SRC}^*]^n + k_{m+}^n} - k_-\frac{\Gamma_-[\mathrm{SRC}^*]}{[\mathrm{SRC}^*] + k_{m-}} + k_-^{\mathrm{Is}}[\mathrm{SRC}_\mathrm{I}^*] - k_+^{\mathrm{Is}}[\mathrm{SRC}^*][I_{\mathrm{SRC}}], \tag{3.20a}$$

$$\frac{d}{dt}[\mathrm{SRC}_\mathrm{I}] = k_+^{\mathrm{Is}}[\mathrm{SRC}][I_{\mathrm{SRC}}] - k_-^{\mathrm{Is}}[\mathrm{SRC}_\mathrm{I}] + k_-[\mathrm{SRC}_\mathrm{I}^*] - k_+[\mathrm{SRC}_\mathrm{I}][L-R], \tag{3.20b}$$

$$\frac{d}{dt}[\mathrm{pMEK}_{\mathrm{total}}] = \Gamma_M\frac{[\mathrm{SRC}^*]^N}{[\mathrm{SRC}^*]^N + k^N} - \lambda[\mathrm{pMEK}_{\mathrm{total}}], \tag{3.20c}$$

$$\frac{d}{dt}[\mathrm{MEK}_\mathrm{I}] = k_+^{\mathrm{Im}}[\mathrm{pMEK}][I_{\mathrm{MEK}}] - k_-^{\mathrm{Im}}[\mathrm{MEK}_\mathrm{I}], \tag{3.20d}$$

$$\frac{d}{dt}[\mathrm{ppERK}] = \Gamma_E[\mathrm{pMEK}] - \lambda[\mathrm{ppERK}]. \tag{3.20e}$$

In this model all equilibrium rate parameters are represented using $k_{+,-}^{\theta}$ notation, where $(+,-)$ represent the forward and reverse rates respectively and θ provides

the corresponding reaction identity. Specifically, no superscript represents the $L - R$ and SRC interaction, 'Is' superscript is the SRC and I_{SRC} interaction, and lastly 'Im' superscript is the pMEK and I_{MEK} interaction. The feedbacks are parameterized by the half effective concentration constants for the forward reaction k_{m+} and the reverse reaction k_{m-}, the hill coefficient of the positive feedback n, and the positive (Γ_+) and negative (Γ_-) feedback strengths. Lastly, all dephosphorylation rates, λ, are identical.

Analysis of the equations are broken into two subnetworks - first the dynamics of [SRC*], followed by those of [MEK] and [ppERK]. To analyze the dynamics of [SRC*] we first define the constraints originating from mass conservation,

$$[\text{SRC}_{\text{total}}] = [\text{SRC}] + [\text{SRC}^*] + [\text{SRC}_{\text{I}}] + [\text{SRC}^*_{\text{I}}], \tag{3.21a}$$

$$[L - R_{\text{total}}] = [L - R] + [\text{SRC}^*] + [\text{SRC}^*_{\text{I}}]. \tag{3.21b}$$

In our experiments we administer saturating doses of antigen bound to peptide MHC on antigen presenting cells, $[L - R_{\text{total}}] \gg [\text{SRC}_{\text{total}}]$, making $[L - R_{\text{total}}] \approx [L - R]$ a reasonable approximation. Furthermore, we apply a quasi-steady state approximation in which all equilibrium reactions are much faster than enzymatic reactions. Together we see that

$$[\text{SRC}] = \frac{[\text{SRC}_{\text{total}}]}{1 + \frac{[I_{\text{SRC}}]}{k_I^{\text{src}}}} - [\text{SRC}^*] \tag{3.22}$$

with $k_I^{\text{src}} = k_-^{\text{Is}}/k_+^{\text{Is}}$.

We continue by normalizing our data by k_{m+} allowing us to rewrite the dynamics of [SRC*] by the unitless variable x. Specifically we take $x = [\text{SRC}^*]/k_{m+}$, $\alpha = k_{m-}/k_{m+}$, $s_t = [\text{SRC}_{\text{total}}]/k_{m+}$, $\Gamma_-/k_{m+} = \Gamma'_-$, and $\mathcal{I}_s = [I_{\text{SRC}}]/k_I^{\text{src}}$ to obtain

$$\frac{1}{\Gamma'_- k_-}\frac{d}{dt}x = \frac{\Gamma_+[L-R]}{\Gamma'_- k_d}\left[\frac{1}{\Gamma_+} + \frac{x^n}{x^n+1}\right]\left[\frac{s_t}{1+\mathcal{I}_s} - x\right] - \left[\frac{1}{\Gamma'_-} + \frac{1}{x+\alpha}\right]x. \tag{3.23}$$

Throughout our experiments we do not change the quantity of antigen or antigen presenting cells allowing us to group $\frac{\Gamma_+[L-R]}{\Gamma'_- k_d}$ into a single dimensionless parameter A_1. Furthermore, we take the feedback amplitudes to be much larger than 1, resulting in $\frac{1}{\Gamma_+}, \frac{1}{\Gamma'_-} \to 0$ and a simpler expression for x,

$$\frac{d}{d\tau}x \approx A_1 \underbrace{\frac{x^n}{x^n+1}}_{\text{Positive Feedback}} \overbrace{\left[\frac{s_t}{1+\mathcal{I}_s} - x\right]}^{\text{Mass conservation}} - \underbrace{\frac{x}{x+\alpha}}_{\text{Negative Feedback}} . \tag{3.24}$$

This simplified equation of the dimensionless variable x provides a function for analyzing the topology dependent response of our experimental system to SRC inhibition. To find the values of x at the fixed points of the system we numerically evaluated the stationary solution to Equation 3.24.

The second subnetwork of our system represents the dynamics of [pMEK] and [ppERK]. The MEK inhibitor binds to free pMEK, which results in an inactive complex MEK_I. As a result the total pMEK (pMEK_total) produced by SRC* exists in two states. To link the dynamic equation governing [pMEK_total] (Eq 3.20c) with the uninhibited pMEK species - the species that produces our quantity of interest ([ppERK]) - we write the following mass conservation equation [pMEK_total] = [pMEK] + [MEK_I]. In addition, we assume that the kinetics of the drug binding to pMEK are much faster than the enzymatic reactions. Incorporating the constraint and approximation we see that [pMEK_total] = [pMEK]$(1 + k_I^{\text{mek}})$, allowing us to proceed by writing an expression for the dynamics of [pMEK] with respect to Eq 3.20c as

$$\frac{d}{dt}[\text{pMEK}] = \frac{\Gamma_M}{1+\mathcal{I}_m}\frac{[\text{SRC}^*]^N}{[\text{SRC}^*]^N + k^N} - \lambda[\text{pMEK}]. \tag{3.25}$$

We continue by rewriting the expression with respect to the unitless variables

$$\frac{d}{d\tau}m = \frac{\gamma_m}{1+\mathcal{I}_m}\frac{x^N}{x^N + k_x^N} - \Lambda m. \tag{3.26}$$

Where $\gamma_m = \frac{\Gamma_M}{k_{m+}k_-\Gamma'_-}$, $\Lambda = \frac{\lambda}{k_-\Gamma'_-}$, $x = [\text{SRC}^*]/k_{m+}$, and $m = [\text{pMEK}]/k_{m+}$. Similar to the dynamic equations for unitless pMEK the dynamics of [ppERK] in Eq 3.20e can be rearranged,

$$\frac{d}{d\tau}y = \gamma_e m(\tau) - \Lambda y, \tag{3.27}$$

where $\gamma_e = \frac{\Gamma_e}{k_-\Gamma'_-}$ and $y = [\text{ppERK}]/k_{m+}$.

By solving the differential equation for m we can simplify our system of three equations to two. Two simple dynamic equations is ideal, because the system properties can be easily interpreted in a single two-dimensional phase diagram (Figure 3.12). The m equation can be described by the following integral equation

$$m(\tau) = \frac{\gamma_m}{\Lambda}\frac{1}{1+\mathcal{I}_s}\left[\underbrace{\frac{x^N(\tau_o)}{x^N(\tau_o)+k_x^N}e^{-\Lambda(\tau-\tau_o)}}_{A} + \underbrace{\Lambda\int_{s>\tau_o}^{\tau}\frac{x^N(s)}{x^N(s)+k_x^N}e^{-\Lambda(\tau-s)}ds}_{B}\right]. \tag{3.28}$$

The integral equation representation provides transparency to the systems behavior at the initial conditions, $\tau = \tau_o$ (A), and the resulting dynamics, $\tau > \tau_o$ (B). Because we are studying the dependence of the instantaneous flux for different $x(\tau_o)$ and $y(\tau_o)$ we set the B term to zero. Substituting for m and $\gamma_m\gamma_e/\Lambda^2 = A_0$, the initial fluxes are described as

$$\boxed{\begin{aligned}\left.\frac{d}{d\tau}x(\tau)\right|_{\tau=\tau_o} &= A_1\frac{x(\tau_o)^n}{x(\tau_o)^n+1}\left[\frac{s_t}{1+\mathcal{I}_s}-x(\tau_o)\right]-\frac{x(\tau_o)}{x(\tau_o)+\alpha}, &\text{(3.29a)}\\ \left.\frac{d}{d\tau}y(\tau)\right|_{\tau=\tau_o} &= \frac{\Lambda A_0}{1+\mathcal{I}_m}\frac{x(\tau_o)^N}{x(\tau_o)^N+k_x^N}-\Lambda y(\tau_o). &\text{(3.29b)}\end{aligned}}$$

The Equations 3.29 represent the instantaneous flux and are used for the stream plots of the [ppERK] versus [SRC*] (y and x) phase planes in Figure 3.12. We proceed by finding the properties of our equations that are experimentally measured

- namely the locations and the respective stabilities of the stationary values,

$$\boxed{\begin{aligned} 0 &= A_1 \frac{x^n}{x^n+1}\left[\frac{s_t}{1+\mathcal{I}_s} - x\right] - \frac{x}{x+\alpha}, && \text{(3.30a)} \\ y &= \frac{A_0}{1+\mathcal{I}_m}\frac{x^N}{x^N + k_x^N}. && \text{(3.30b)} \end{aligned}}$$

We evaluated the fixed points of x and their corresponding stability numerically in accordance to Eq 3.30a. The stability of each fixed point was assessed by numerical evaluating the direction of the flux, positive or negative, for small displacements about the respective fixed point. Given our chosen parameter set (Table 3.2), we found our system to elicit bistable behavior [78].

Table 3.2: **TCR model parameters.** Parameters used for all TCR modeling figures.

Parameter	Interpretation	Assignment
A_1	Effective x forward rate constant	1.75
s_t	Quantity of dimensionless [$\text{SRC}_{\text{total}}$]	2.5
n	steepness of positive feedback	3
α	negative feedback sensitivity / positive feedback sensitivity	0.1
A_0	Effective y production constant	1
k_x	sensitivity of y to changes in x	0.5
N	hill coefficient of MAPK	4.
Λ	Effective y degradation constant	1.

To develop intuition regarding the properties of our competing feedback model we constructed phase diagrams of active SRC that demonstrate how varied quantities of SRC* map to ERK phosphorylation. SRC* originally accumulates upon engagement of the activating ligand with its kinase-bearing receptor. This subsequently activates both positive and negative feedback on further accumulation of SRC*. The dynamics of accumulation of SRC* can be summarized in a phase diagram (Figure 3.12B), that illustrates the influence of both feedbacks. The model parameters are set so that the negative flux (or time derivative of SRC*) rises and saturates at lower levels of SRC* than the positive flux (Table 3.2). This staggering of the positive and negative fluxes as a function of SRC* causes them be

equal at three points in the phase diagram, *i.e.* there are three stationary or fixed points. By plotting the net flux as a function of the active complex SRC*, one can assess the stability of the fixed points. The dynamics are such that SRC* always converges to the extreme points $\text{SRC}^*_{\text{low}}$ and $\text{SRC}^*_{\text{high}}$ (stable fixed points), while diverging from the center point $\text{SRC}^*_{\text{med}}$ (unstable fixed point). This encapsulates the bistability in SRC* formation. We then model ERK activation by assuming that the active complex SRC* enzymatically phosphorylates MEK, which then proportionally phosphorylates ERK (Figure 3.12C). In Figure 3.12D, we represent the dynamic trajectory of this signaling pathway for varied initial conditions: such flow diagram illustrates the stability of the low and high states in the (SRC*,ppERK) plane, the instability of the intermediate point. Overall, our coarse grained model of ERK activation upon ligand engagement generates two stable fixed points corresponding to either zero or maximum ppERK, consistently with our experimental results.

We then tested how our coarse-grained model would predict ERK response to drug inhibition. Application of the MEK inhibitor to our model (Figure 3.12E) supports our experimental observations. Inhibition of MEK does not influence the bistability of the activated kinases SRC* (Figure 3.12F), and does not itself support nonlinearities. Titrating the MEK inhibitor shows continuous reduction in the amplitude of ppERK response (Figure 3.12G). Figure 3.12H explicitly shows that the dynamic properties supporting the bistability in ppERK are preserved in the presence of the MEK inhibitor. Our model is validated with the experimental observation that the MEK inhibitor only reduces the average quantity of ppERK of the activated cells *i.e.* it inhibits the ERK pathway in an analog manner.

SRC is the kinase that is critical to establish the bistability of the active complex

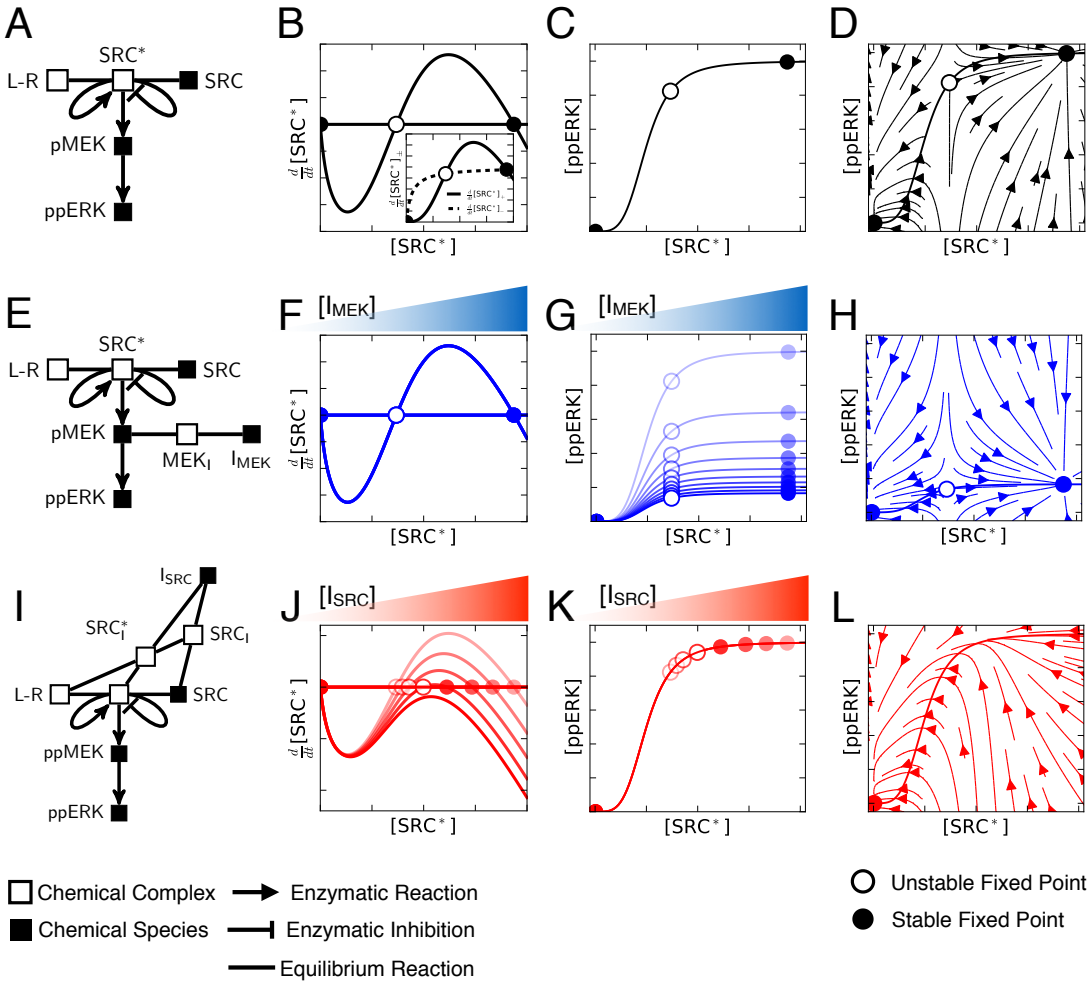

Figure 3.12: **Subnetwork context of the targeted enzyme determines response to inhibition.** (A) Model diagram of signaling network. (B) Phase plane of SRC*, inset shows the behavior of both the positive and negative model fluxes. (C) Functional response of ppERK to changes in SRC* abundance. (D) Instantaneous reaction velocities given ordered pairs of (SRC*, ppERK) shows the dynamic behavior of the model system. (E) Model diagram for MEK inhibition of signaling. (F,G) Analogous to representations of model behavior as (B,C) but for different doses of MEK inhibitor. (H) Instantaneous reaction velocities for maximal MEK inhibition. (I) Model diagram for SRC inhibition. (J,K) Analogous representations of SRC inhibition as (F,G) for MEK inhibition. (L) Instantaneous reaction velocities for maximal SRC inhibition show that the dynamics support a single fixed point at (SRC*, ppERK) = (0,0).

SRC*, resulting in a signaling context fundamentally distinct to that of pMEK. Inhibition of SRC reduces the positive flux generating SRC* (Figure 3.12I), and consequently reduces SRC* at the high fixed point. Increasing the dose of the SRC inhibitor decreases SRC* until, at a critical dose, the high fixed point and the unstable fixed point annihilate one another (Figure 3.12J). Doses of SRC inhibitor greater than the critical dose result in a system with a single fixed point, SRC*= 0 (Figure 3.12J). Despite continuous reduction of the high stable fixed point with increased dosage of SRC inhibitor, quantities of ppERK remain negligibly changed until the inhibitor is greater than the critical dose (Figure 3.12K). For doses of SRC inhibitor beyond the critical dose the signaling network only supports a single quantity ppERK at zero (Figure 3.12L). Hence, SRC inhibition results in a binary output that is identical to that observed in the data: our model is consistent with the digital nature of Dasatinib as a SRC inhibitor.

Our model assumed that interactions of molecular inhibitors with their target enzyme all act as noncompetitive inhibitors with respect to the biochemical substrate. Yet, this biochemically identical mechanisms of drug inhibition is sufficient to account for the two distinct modes of inhibition of ERK in our experimental findings. Thus, our results and theoretical model demonstrate that enzymatic context is an essential component for parameterizing inhibitor function in living cells.

3.2.3 Variability in protein abundance diversifies sensitivity of signaling pathway to targeted inhibition

Using our mechanistic model, we sought to explore how the endogenous variability of SRC abundance would diversify the response of individual cells to inhibition. Our model predicted that the effective quantity of SRC determines whether the (SRC, ppERK) phase diagram has a single or three fixed points - as a result it represents a bifurcation parameter. By analogy, endogenous variation of SRC positions cells above and below the critical quantity of SRC required for bistable signaling (Figure 3.13A). We tested this model generated hypothesis by measuring CD8 (as a proxy for the abundance of SRC) and ppERK of activated T cells. Indeed, in T cells LCK, a Tyrosine kinase is recruited by CD8 for receptor signaling, therefore we assumed that CD8 abundance could act as a proxy for the effective abundance of SRC in individual cells. Indeed, measuring CD8 for a single dose of SRC inhibitor shows that cells with elevated quantities of CD8 are more likely to have ppERK signal (Figure 3.13B), a result that is consistent with previous experimental and theoretical observations [31]

Extending this observation, our model suggests that the variability of CD8 expression in single cells is sufficient to generate disparate sensitivities to drug inhibition. The bifurcation diagram for each drug dose, Figure 3.13C, shows that the minimum quantity of SRC sufficient for the bistability, SRC_c, increases with drug dose. Consequently, a cell with a higher quantity of SRC will be more tolerant to inhibition because of simple dosing of the available SRC (Figure 3.13D) - requiring higher inhibitor dosage to experience any reduction in signaling.

Indeed, CD8 abundance per cell does not perfectly predict whether a cell is

capable of signaling (Figure 3.13B). As a result we estimated the half effective abundance of CD8 (a.u.) as an effective measure of CD8's contribution to tuning the sensitivity of individual cells to drug. We call this quantity, $CD8_c$, and estimate it by fitting the fraction of cells signaling to a hill function,

$$f(\mathrm{CD8}; \mathrm{CD8}_{ci}, \vec{\theta}_i) = \theta_{0i} \frac{\mathrm{CD8}^{\theta_{1i}}}{\mathrm{CD8}^{\theta_{1i}} + \mathrm{CD8}_{ci}^{\theta_{1i}}} + \theta_{2i}, \tag{3.31}$$

dependent on CD8 abundance for each, i, dose of inhibitor. In Equation 3.31 θ_0 represents the maximum fraction of cells that can signal, θ_1 represents the sharpness of the CD8 dependency, and lastly θ_2 is the minimum fraction of cells that can signal. The data were fit to Equation 3.31 by minimizing the sum-square residuals with the constraint that $\theta_2 \geq 0$ - no other constraints were required. The output of our procedure is summarized in Figure 3.13E which shows the amplitude (θ_{0i}) normalized data and the corresponding hill fits. We applied this procedure to both the MEK inhibitor and SRC inhibitor data, and confirmed our predictions that the MEK inhibitor reduced ERK activation independently of the levels of CD8 while higher concentrations of SRC inhibitor correlated with a higher $CD8_c$ (Figures 3.13F).

3.2.4 Diverse signaling responses to drug inhibition on short time-scales translate into diverse cell proliferative response on long time-scales.

Upon phosphorylation, ERK migrates from the cytosol to the cell nucleus where it induces the expression of the immediate and early genes (IEGs, *e.g.* cFOS). IEGs constitute a set of genes that facilitate cell cycle entry and cell division [57], hence

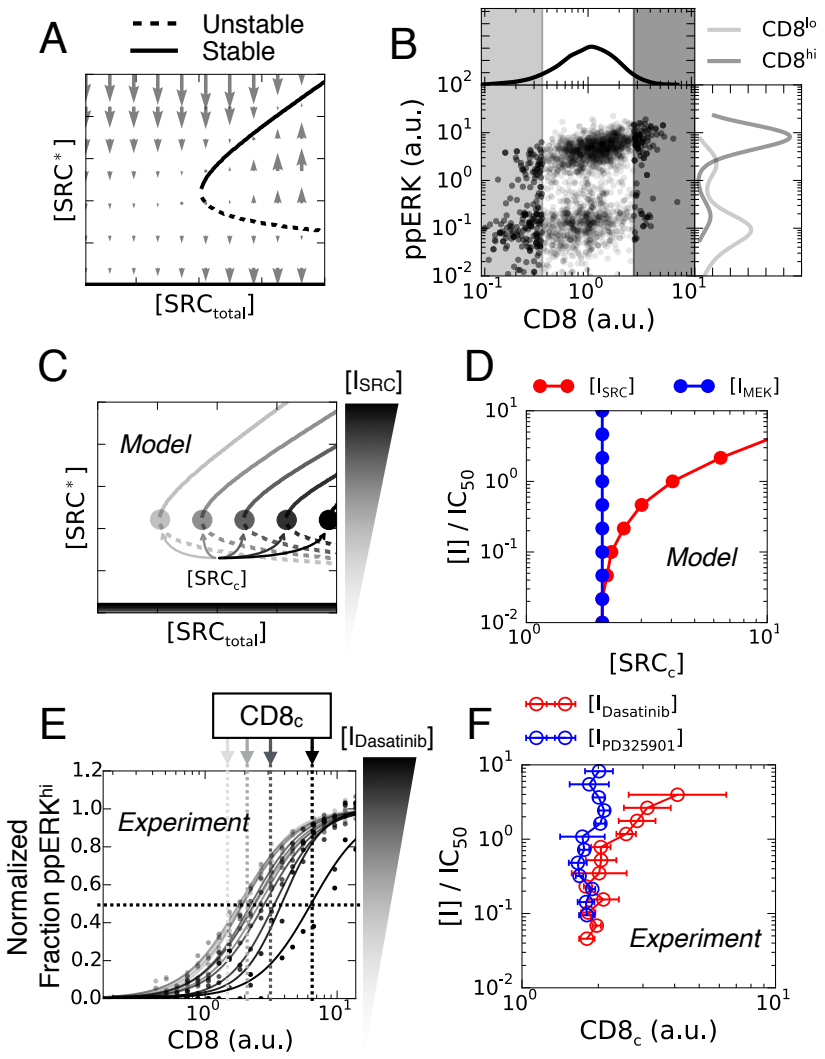

Figure 3.13: **Protein variability tunes sensitivity of cells to inhibition.** (A) Mechanistic model reveals that the endogenous expression of SRC functions as a bifurcation parameter. (B) Flow cytometry measurements of T cells concomitantly labeled for CD8 (a proxy for SRC) and ppERK (pT202, pY204) shows signaling dependence to endogenous expression of CD8. (C) Model predictions of SRC_c dependence with SRC inhibitor dose. (D) Model predicts scaling of SRC_c with increasing doses of a SRC inhibitor and not a MEK inhibitor. (E) CCVA of T cells concomitantly labeled for CD8 and ppERK (pT202, pY204) treated with various doses of the SRC inhibitor Dasatinib. (F) Quantification experimental data of the half effective abundance of CD8 ($CD8_c$) required for ppERK activation in T cells treated with either MEK inhibitor (PD325901) or SRC inhibitor (Dasatinib).

one functional readout of ERK activation is the overall cell proliferation response. Based on our signaling results (Figure 3.10), we conjectured that intermediate levels of pERK from titrating a MEK inhibitor may induce slower induction of IEGs, and as a result would increase the time to cell division. In contrast, titrating a SRC inhibitor would change the fraction of cells entering the cell cycle as its does with the fraction of cells maximally inducing ppERK, without affecting the overall cell division among activated cells (cf Figure 3.10) [3].

We tested this hypothesis by quantifying the cell proliferation of T cells after 48 hours of *in vitro* culture under concomitant antigen stimulation and drug exposure. We monitored cell activation and division by measuring cell size, the levels of the CD8 coreceptor on surface of cells, and the fluorescence of T cells that were tagged before activation with the amine-reactive fluorescent dye CTV or CFSE (the dye gets diluted by two at each cell division) by flow cytometry. Upon activation cells increase both their size and CD8 expression, providing a clear criteria separating active and inactive cell populations, whose numbers can be quantified as $\mathrm{N}^{\pm}$ respectively (Figure 3.14A). We then analyzed the number of cells ($\mathrm{N}_i^{\pm}$ for $i = \{0, 1, 2, 3\}$) undergoing i divisions as measured by CTV dilution (Figure 3.14B). We computed the average number of divisions as,

$$\left\langle \mathrm{Divisions}^{\pm} \right\rangle = \frac{1}{N_{total}^{\pm}} \sum_{n=0}^{n_{max}} \frac{N_n^{\pm}}{2^n} n, \tag{3.32}$$

where N_n is the number of cells in the n division gate, and

$$N_{total}^{\pm} = \sum_{n=0}^{n_{max}} \frac{N_n^{\pm}}{2^n}. \tag{3.33}$$

Meanwhile, the fraction of activated cells are computed as,

$$\text{Fraction Active} = \frac{\left\langle \text{Divisions}^{+} \right\rangle}{\left\langle \text{Divisions}^{\text{all}} \right\rangle} \tag{3.34}$$

$$= \frac{N_{total}^{\text{all}}}{N_{total}^{+}} \frac{\sum_n \frac{N_n^+}{2^n} n}{\sum_n \frac{N_n^{\text{all}}}{2^n} n} \tag{3.35}$$

$$= \text{Active Fraction} \underbrace{\frac{\sum_n n \frac{N_n^+}{2^n}}{\sum_n n \frac{N_n^{\text{all}}}{2^n}}}_{\text{Gate correction}} . \tag{3.36}$$

We have found that computing the active fraction in this way increases the robustness of the analysis. For many doses, simple gating of active and inactive populations resulted in the bifurcation of what appeared as a single distribution of cells. By adding the extra information that only the active cells proliferate and divide, we can correct for the error in estimating the active fraction that is introduced by the simple gating analysis.

Representing the data with respect to (mean divisions, active fraction) clearly demonstrates how the disparate modes of inhibition from signal transduction map to the proliferative timescale. Furthermore, we see that the mode of inhibition depends solely on the inhibitor target species and not the identity of chemical inhibitor applied (Figure 3.14C). Explicitly, we see that dosing a MEK inhibitor reduces the average number of divisions among activated cells. While the dominant feature of the SRC inhibition is the distinct reduction of activated cells. Indeed, the fraction of activated cells is the dominant but not the exclusive feature observed in our data, intermediate dosages of SRC inhibition does reduce the mean divisions. While our model does not produce insight to this phenomena, it is not surprising, because of the unaccounted signaling transduction pathways dependent on TCR activation, *e.g.* PI3K and AKT [87].

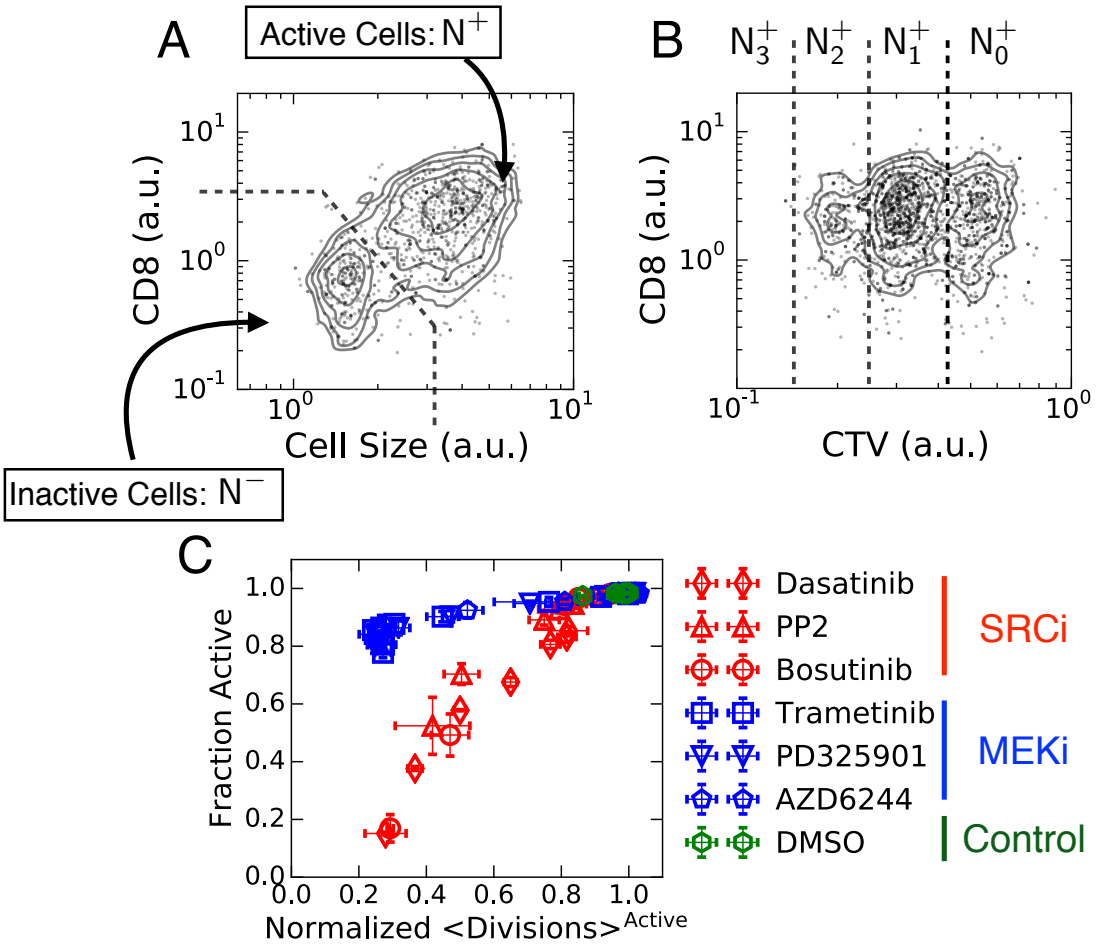

Figure 3.14: **Short time scale modes of inhibition translate to long timescale proliferative response.** (A) Identification of activated T cells after 48 hours of concomitant treatment with antigen and the respective inhibitor measured by flow cytometry. Upon activation, cells increase size, measured by forward scatter area (FSC-A), and up-regulate surface protein CD8. (B) Cell trace violet (CTV) dilution identifies active cells that have divided n times. (C) Quantification of fraction of active cells and mean number of divisions for each chemical inhibitor.

3.3 Summary

In this study we use a combined theoretical and experimental approach to extend the mechanistic understanding of chemical inhibitors to living cells. We use single cell phospho-profiling and Cell to Cell Variability Analysis (CCVA) [22] to perform detailed biochemical characterization of small molecule chemical inhibitors - to date such detail has been limited to *in vitro* enzymatic assays. We also uncovered a generalizable mechanism in which targeted enzyme inhibition manifests to unique patterns of inhibition, digital and analog. Lastly, we probed the biological significance of these results from short-timescale signaling behavior to unique modes of inhibition of cellular proliferation.

CHAPTER 4

CONCLUDING REMARKS

In this work we presented results demonstrating the signatures of biological variability at the molecular scale. In doing so, we developed tools to study the transmission of fluctuations in the abundance of biochemical species throughout gene regulatory networks and biological signaling systems. We have also shown how single cell measurements elucidate complex phenomena of signaling pathways when perturbed by small molecule chemical inhibitors. Moreover, we showed how the endogenous variability of protein can diversify the response of individual cells to chemical inhibition. This introduced diversity provides sufficient data to perform detailed biochemical titration analysis of chemical inhibitors in living systems. Taken together, we have made progress in understanding diversity at the molecular scale, and may facilitate understanding the molecular origins of non-genetic phenotypic variability observed in biological systems.

4.1 Propagation of biochemical fluctuations

Biological systems utilize networks of biochemical interactions to process information regarding their environment and illicit appropriate physiological responses. Indeed, these systems are constrained by the stochastic properties of their molecular constituents, namely stochasticity of the chemical reactions. Moreover, biochemical systems are complex, representing an aggregate of stochastic processes that manifest as fluctuations of molecular components that exceed those predicted by one step stochastic processes [82, 73, 33]. Consequently, appropriate descriptions of the fluctuations in the abundance of molecular components must reflect these long tailed distributions. To solve this problem, we developed a chemical

Langevin equation of the logarithmic concentrations of molecular species in accordance with single cell experimental observations (Figures 2.2,2.6). With this equation we derive a Lyapunov equation (Eq 2.29) to study the transmission of biological fluctuations of molecular components theoretically and validate these results in two experimental model systems: a synthetic *E. coli* gene regulatory network and mammalian phospho-signaling.

A premise of our analysis is that fluctuations in the abundance of biochemical species are approximately log-normal. To account for these log-normal fluctuations we derived a chemical Langevin equation of the logarithmic biochemical concentrations by considering fluctuations of rate parameters and asserting that the magnitude of the fluctuations are independent of network activity. Indeed, our consideration of parameter fluctuations is purely empirical. However, we have two sources of experimental evidence of this description. First, we are able to describe the propagation of fluctuations in biochemical components by analyzing the resulting covariance and variance plots in *E. coli* (Figure 2.11) and T lymphocytes (Figure 2.13). Moreover, in each set of measurements, the covariance of maximal and minimal network activity were identical. This suggests that the fluctuations in our experiments are not dominated by extrinsic sources. Indeed, extrinsic sources of variability would manifest in covariances that differ at maximal and minimum stimulation. In addition, we used our model to estimate the magnitude of internal noise as it depends on the activity of the pathway. We report that the internal logarithmic fluctuations in T lymphocyte phospho-signaling is invariant to the pathway activity (Figure 2.16). Together, these results justify our analysis, yet more detailed investigation to the precise origins are needed (*e.g.* single cell temporal measurements of each process).

The utility of our approach comes from our derived Lyapunov equation (Equation 2.29). This equation provides a description of the covariance matrix as a function of both the network connectivity and the internal noise. From this equation, we showed how different network structures produce unique functional relationships between covariance matrix elements. Analysis of these elements with changing connectivity strengths provides signatures of the network structure. Indeed, when comparing a pair of interacting biochemical species, the downstream molecule will sum its internal fluctuations and those propagated from its upstream partner. This intuition is captured by plotting the data in the covariance and variance plane. For example, in the isolated two node interaction the down stream molecule increases its variance as the square of the covariance. Meanwhile the variance of the upstream component is invariant to changes in the covariance (Figure 2.7). In addition, the presence of a third unmeasured quantity can be revealed by deviations from the ideal two node behavior (Figure 2.8,2.9). These properties are advantageous for studying the structure of signaling systems, indeed these functional relationships are useful for inference tasks.

This analysis also provides information of the signaling networks parameterization. Stimulating MAPK with PMA activates a cascade of biochemical reactions that ultimately lead to the phosphorylation of MEK and ERK (Figure 2.3). Yet, while the biology suggests that the cascade model should be superior, our data are well approximated by the isolated two node system. Figure 2.10 demonstrates the sensitivity of the functional relationships between covariance elements to changes in the parameters of the cascade network. Interestingly, the biology of MAPK signaling has selected for a parameterization that eliminates loops in the covariance and variance plane. This suggests, that these systems make concerted efforts to appear as an isolated two-node system [83]. It is intriguing that cells can tune their

biochemical parameters to control for transmitted fluctuations instead of relying upon negative feedback regulation alone.

Our results have shown that the combination of a simple single cell assay and analysis of covariance elements provides insights to the structure of biochemical networks. The functional relationships between covariance elements provide insights to the direction of biochemical reactions, identify the existence of unmeasured components, and reveal novel network parameterization.

4.2 Variable responses of individual cells to chemical inhibition

In this study we use a combined theoretical and experimental approach to extend the mechanistic understanding of chemical inhibitors to living cells. We use single cell phospho-profiling and Cell to Cell Variability Analysis (CCVA) [22] to perform detailed biochemical characterization of small molecule chemical inhibitors - to date such detail has been limited to *in vitro* enzymatic assays. We also uncovered a generalizable mechanism in which targeted enzyme inhibition manifests to unique patterns of inhibition, digital vs. analog. Lastly, we probed the biological significance of these results from short-timescale signaling behavior to unique modes of inhibition of cellular proliferation.

Using single cell phospho-profiling and CCVA we were able to perform detailed and mechanistic characterization of cellular responses to targeted inhibition in primary cells. Specifically, we showed how to utilize CCVA to mechanistically characterize the biochemical interaction between the enzyme target and the inhibitor.

We found that AZD1480 is a potent noncompetitive inhibitor JAK-STAT signaling in IL-2 stimulated primary T lymphocytes (Figure 3.8). Remarkably, our measurements highlight the promiscuity of AZD1480 when administered to primary cells - IL2 receptor associates with JAK1 and JAK3 and not the 'specific' enzyme target of the drug, JAK2 [55]. In addition, we demonstrated how the organization of reactions in biochemical networks determines the cellular response, analog or digital, to inhibition (Figures 3.10,3.12). Although, Albeck *et al* also found that inhibition of different enzymes manifest to digital or analog signaling responses [3], we uniquely show how these disparate responses to inhibition are attributable to the context of the targeted enzyme. Indeed, the dynamic properties associated with the targeted enzyme's location in the larger biochemical network determines the networks response to inhibition (Figure 3.12). Furthermore, we showed how our short timescale signaling behavior translates to novel long timescale proliferative response to inhibition (Figure 3.14). As a result, our methodology extends *in vitro* characterization of inhibitor specificity and mechanism to primary single cells and provides insights to nonlinear signaling responses from the biochemical network structure.

Building upon our initial findings, we extended our mechanistic models and used CCVA to demonstrate how cells utilize the endogenous variability of protein abundance to generate disparate responses to singular perturbations. In context to inhibition, we found that variation in enzyme substrate (STAT5) abundance established diverse signaling amplitudes and varied the sensitivity of single cells to inhibition (Figure 3.1,3.8). We then extended our mechanistic model of SRC inhibition and found that the variability of SRC expression operates on cells as a bifurcation parameter. As a result, cells that had elevated abundance of SRC are more tolerant to inhibition (Figure 3.13). The extent in which these mechanisms of

diversity provide resilience of populations to inhibition at longer timescales remains an open question. However, the numerous examples of biological systems that utilize protein abundance to generate phenotypic variability [28, 79, 76, 17, 41], and the abundance of single cell observations showing heterogeneous responses of cells to inhibition [75, 12, 14] suggest our findings are of practical importance.

By using our combined approach of CCVA and development of mechanistic models to characterize drugs we were able to attribute single cell observations to fundamental chemical and biological processes. We hope that this methodology facilitates the extension of *in vitro* kinase assays to cellular systems and motivates a transition from phenomenological characterization of single cell responses to drugs into mechanistic understanding.

4.3 Future work

The goal of this work was to understand the properties of biological fluctuations in the abundance of molecules and the variable responses of individual cells to chemical inhibition. We have provided tools and methodologies to study single cell processes, however our analysis has been constrained to snapshots of the underlying phenomena. These studies would be enhanced by using single cell temporal measurements of the respective processes.

Temporal measurements of pMEK and ppERK activity would allow us to better define the properties of the network. Currently, our fluctuation analysis assumes that the fluctuations in molecular abundance are manifestations of stochasticity in the chemical reactions and parameters, yet we do not have proof of this mechanism. To prove such a mechanism we need to assess three components that are in

Equation 2.28: the magnitude of the correlations in the abundance of molecules, the timescale of the observed correlations, and lastly the presumed ergodicity of the system. Our single cell assays of signal transduction preclude the repeated observations of biochemical signaling in individual cells. While this does not prevent estimation to the magnitude of biochemical fluctuations, it forbids the proper assessment of the correlation timescale - by computing the autocorrelation and cross-correlation functions - and proof of ergodicity. As a result, temporal measurements of signaling in single cells is the most pressing next step in this research.

Furthermore, repeated cell measurements would facilitate our mechanistic understanding of chemical inhibitors. Our current model of signaling suggests that the endogenous variability of protein tunes the responsiveness of cells to SRC inhibition. We were not able to extend these findings to the proliferative timescale, because we could not track individual cells over time. Concomitant monitoring of single cell ppERK signaling and CD8 in time would provide the necessary measurements to demonstrate our prediction.

Pioneering work by Albeck *et al.* [3] used FRET reporters of ERK signaling to monitor signaling over time [47, 5]. Strikingly, they observed that this pathway produced stochastic pulses that changed frequency with dose of Epidermal Growth Factor (EGF), a potent activator of MAPK signaling in the human mammary cell line, MCF10A. The most striking and under-appreciated finding was that the patterns of activity of each cell were unique, and that they were able to switch in time. For example, some cells would be activated for several hours, then spontaneously switch to limit cycle oscillations, and back. The underlying mechanisms of the observed diversity were not investigated by the authors. In context to our assumed parameter fluctuations and the nonlinearities in the MAPK pathway [67],

this system can easily sample from the observed dynamics phenotypes. It would be fascinating if biological system purposefully used fluctuations in parameters (*e.g.* protein concentration per cell) to sample dynamic signaling phenotypes as a mechanism to maintain physiological viability.

BIBLIOGRAPHY

[1] M Acar, JT Mettetal, and A van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics*, 40(4):471–475, 2008.

[2] John G. Albeck, John M. Burke, Bree B. Aldridge, Mingsheng Zhang, Douglas A. Lauffenburger, and Peter K. Sorger. Quantitative Analysis of Pathways Controlling Extrinsic Apoptosis in Single Cells. *Molecular Cell*, 30(1):11 – 25, 2008.

[3] John G. Albeck, GordonB. Mills, and JoanS. Brugge. Frequency-modulated pulses of erk activity transmit quantitative proliferation signals. *Molecular Cell*, 49(2):249 – 261, 2013.

[4] Grégoire Altan-Bonnet and Ronald N Germain. Modeling T Cell Antigen Discrimination Based on Feedback Control of Digital ERK Responses. *PLoS Biol*, 3(11):e356, 10 2005.

[5] Kazuhiro Aoki, Yuka Kumagai, Atsuro Sakurai, Naoki Komatsu, Yoshihisa Fujita, Clara Shionyu, and Michiyuki Matsuda. Stochastic ERK Activation Induced by Noise and Cell-to-Cell Propagation Regulates Cell Density-Dependent Proliferation. *Molecular Cell*, 52(4):529 – 540, 2013.

[6] Adam Arkin, John Ross, and Harley H McAdams. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected *Escherichia coli* Cells. *Genetics Society of America*, 149:1633–1648, 1998.

[7] Nathalie Q. Balaban, Jack Merrin, Remy Chait, Lukasz Kowalik, and Stanislas Leibler. Bacterial Persistence as a Phenotypic Switch. *Science*, 305(5690):1622–1625, 2004.

[8] Marcus Bantscheff, Dirk Eberhard, Yann Abraham, Sonja Bastuck, Markus Boesche, Scott Hobson, Toby Mathieson, Jessica Perrin, Manfred Raida, Christina Rau, Valerie Reader, Gavain Sweetman, Andreas Bauer, Tewis Bouwmeester, Carsten Hopf, Ulrich Kruse, Gitte Neubauer, Nigel Ramsden, Jens Rick, Bernhard Kuster, and Gerard Drewes. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotech*, 25(9):1035–1044, 2007.

[9] Jose Baselga, Larry Norton, Joan Albanell, Young-Mee Kim, and John Mendelsohn. Recombinant Humanized Anti-HER2 Antibody (Herceptin™) Enhances the Antitumor Activity of Paclitaxel and Doxorubicin against HER2/neu Overexpressing Human Breast Cancer Xenografts. *Cancer Research*, 58(13):2825–2831, 1998.

[10] Hubertus J. E. Beaumont, Jenna Gallie, Christian Kost, Gayle C. Ferguson, and Paul B. Rainey. Experimental evolution of bet hedging. *Nature*, 462:90 –93, 2009.

[11] Marcelo Behar, Derren Barken, ShannonL. Werner, and Alexander Hoffmann. The dynamics of signaling as a pharmacological target. *Cell*, 155(2):448 – 461, 2013.

[12] Sean C. Bendall, Erin F. Simonds, Peng Qiu, El-ad D. Amir, Peter O. Krutzik, Rachel Finck, Robert V. Bruggner, Rachel Melamed, Angelica Trejo, Olga I. Ornatsky, Robert S. Balderas, Sylvia K. Plevritis, Karen Sachs, Dana Peer, Scott D. Tanner, and Garry P. Nolan. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science*, 332(6030):687–696, 2011.

[13] Otto G Berg. A Model for the Statistical Fluctuations of Protein Numbers in a Microbial Population. *Journal of Theoretical Biology*, 71:587–603, 1978.

[14] Bodenmiller, Bernd and Zunder, Eli R and Finck, Rachel and Chen, Tiffany J and Savig, Erica S and Bruggner, Robert V and Simonds, Erin F and Bendall, Sean C and Sachs, Karen and Krutzik, Peter O and Nolan, Garry P. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nature Biotech*, 30(9):858–867, 2012.

[15] J Bromberg and JE Darnell. The role of STATs in transcriptional control and their impact on cellular function. *Oncogene*, 19(21):24682473, May 2000.

[16] Long Cai, Nir Friedman, and X. Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(16), 2006.

[17] H.H. Chang, M Hemberg, M. Barahona, D.E. Ingber, and S. Huang. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453(22), 2008.

[18] Paul B. Chapman, Axel Hauschild, Caroline Robert, John B. Haanen, Paolo Ascierto, James Larkin, Reinhard Dummer, Claus Garbe, Alessandro Testori, Michele Maio, David Hogg, Paul Lorigan, Celeste Lebbe, Thomas Jouary, Dirk Schadendorf, Antoni Ribas, Steven J. O'Day, Jeffrey A. Sosman, John M. Kirkwood, Alexander M.M. Eggermont, Brigitte Dreno, Keith Nolop, Jiang Li, Betty Nelson, Jeannie Hou, Richard J. Lee, Keith T. Flaherty, and Grant A. McArthur. Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine*, 364(26):2507–2516, 2011. PMID: 21639808.

[19] Jia-Yun Chen, Jia-Ren Lin, KarleneA. Cimprich, and Tobias Meyer. A Two-

Dimensional ERK-AKT Signaling Code for an NGF-Triggered Cell-Fate Decision. *Molecular Cell*, 45(2):196 – 209, 2012.

[20] Athel Cornish-Bowden. A Simple Mehtod for Determining the Inhibition Constants of Mixed, Uncompetitive and Non-Competitive Inhibitors. *Biochem. J.*, 137:143–144, 1974.

[21] Jesse W Cotari, Guillaume Voisinne, and Grégoire Altan-Bonnet. Diversity training for signal transduction: leveraging cell-to-cell variability to dissect cellular signaling, differentiation and death. *Current Opinion in Biotechnology*, 24(4):760 – 766, 2013. Nanobiotechnology Systems biology.

[22] Jesse W. Cotari, Guillaume Voisinne, Orly Even Dar, Volkan Karabacak, and Grégoire Altan-Bonnet. Cell-to-Cell Variability Analysis Dissects the Plasticity of Signaling of Common gamma Chain Cytokines in T Cells. *Sci. Signal.*, 6(266):ra17, 2013.

[23] Francis Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970.

[24] Jayajit Das, Mary Ho, Julie Zikherman, Christopher Govern, Ming Yang, Arthur Weiss, and Arup K. Chakraborty. Digital signaling and hysteresis characterize RAS activation in lymphoid cells. *Cell*, 136(2):337–351, 2009.

[25] Davis, Mindy I and Hunt, Jeremy P and Herrgard, Sanna and Ciceri, Pietro and Wodicka, Lisa M and Pallares, Gabriel and Hocker, Michael and Treiber, Daniel K and Zarrinkar, Patrick P. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotech*, 29(11):1046–1051, 2011.

[26] F.G. Donnan. Concerning the Applicability of Thermodynamics to the Phenomena of Life. *The Journal of General Physiology*, pages 685–688, 1927.

[27] David P Doupé, Maria P Alcolea, Amit Roshan, Gen Zhang, Allon M Klein, Benjamin D Simons, and Philip H Jones. A Single Progenitor Population Switches Behavior to Maintain and Repair Esophageal Epithelium. *Science*, 337:1091–1093, 2012.

[28] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584):1183–1186, 2002.

[29] Miles A Fabian, William H Biggs III, Daniel K Treiber, Corey E Atteridge, Mihai D Azimioara, Michael G Benedetti, Todd A Carter, Pietro Ciceri, Philip T Edeen, Mark Floyd, Julia M Ford, Margaret Galvin, Jay L Gerlach, Robert M Grotzfeld, Sanna Herrgard, Darren E Insko, Michael A Insko, Andiliy G Lai, Jean-Michel Lelias, Shamal A Mehta, Zdravko V Milanov, Anne Marie Velasco, Lisa M Wodicka, Hitesh K Patel, Patrick P Zarrinkar, and David J Lockhart. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotech*, 23(3):329–336, 2005.

[30] Ofer Feinerman, Garrit Jentsch, Karen E Tkach, Jesse W Coward, Matthew M Hathorn, Michael W Sneddon, Thierry Emonet, Kendall A Smith, and Grégoire Altan-Bonnet. Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Molecular Systems Biology*, 6(1), 2010.

[31] Ofer Feinerman, Joël Veiga, Jeffrey R. Dorfman, Ronald N. Germain, and Grégoire Altan-Bonnet. Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science*, 321(5892):1081–1084, 2008.

[32] Paul François, Guillaume Voisinne, Eric D. Siggia, Grégoire Altan-Bonnet,

and Massimo Vergassola. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism. *Proceedings of the National Academy of Sciences*, 110(10):E888–E897, 2013.

[33] Nir Friedman, Long Cai, and X. Sunney Xie. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.*, 97:168302, Oct 2006.

[34] Daniel T Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[35] Daniel T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[36] Piyush B. Gupta, Tamer T. Onder, Guozhi Jiang, Kai Tao, Charlotte Kuperwasser, Robert A. Weinberg, and Eric S. Lander. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell*, 138(4):645 – 659, 2009.

[37] Douglas Hanahan and RobertA. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646 – 674, 2011.

[38] Katsuhiko Hayashi, Susana M. Chuva de Sousa Lopes, Fuchou Tang, and M. Azim Surani. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell*, 3(4):391 – 401, 2008.

[39] C Y Huang and J E Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences*, 93(19):10078–10083, 1996.

[40] Jonathan M Irish and Randi Hovland and Peter O Krutzik and Omar D Perez and Øystein Bruserud and Bjørn T Gjertsen and Garry P Nolan. Single Cell Profiling of Potentiated Phospho-Protein Networks in Cancer Cells. *Cell*, 118(2):217 – 228, 2004.

[41] Vandana Kalia, Surojit Sarkar, Shruti Subramaniam, W. Nicholas Haining, Kendall A. Smith, and Rafi Ahmed. Prolonged Interleukin-2R Expression on Virus-Specific CD8+ T Cells Favors Terminal-Effector Differentiation In Vivo. *Immunity*, 32(1):91 – 103, 2010.

[42] Tibor Kalmar, Chea Lim, Penelope Hayward, Silvia Muoz-Descalzo, Jennifer Nichols, Jordi Garcia-Ojalvo, and Alfonso Martinez Arias. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol*, 7(7):e1000149, 07 2009.

[43] Hagop Kantarjian, Charles Sawyers, Andreas Hochhaus, Francois Guilhot, Charles Schiffer, Carlo Gambacorti-Passerini, Dietger Niederwieser, Debra Resta, Renaud Capdeville, Ulrike Zoellner, Moshe Talpaz, and Brian Druker. Hematologic and Cytogenetic Responses to Imatinib Mesylate in Chronic Myelogenous Leukemia. *New England Journal of Medicine*, 346(9):645–652, 2002. PMID: 11870241.

[44] Hagop Kantarjian, Neil P. Shah, Andreas Hochhaus, Jorge Cortes, Sandip Shah, Manuel Ayala, Beatriz Moiraghi, Zhixiang Shen, Jiri Mayer, Ricardo Pasquini, Hirohisa Nakamae, Franoise Huguet, Concepcin Boqu, Charles Chuah, Eric Bleickardt, M. Brigid Bradley-Garelik, Chao Zhu, Ted Szatrowski, David Shapiro, and Michele Baccarani. Dasatinib versus Imatinib in Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia. *New England Journal of Medicine*, 362(24):2260–2270, 2010. PMID: 20525995.

[45] Karaman, Mazen W and Herrgard, Sanna and Treiber, Daniel K and Gallant, Paul and Atteridge, Corey E and Campbell, Brian T and Chan, Katrina W and Ciceri, Pietro and Davis, Mindy I and Edeen, Philip T and Faraoni, Raffaella and Floyd, Mark and Hunt, Jeremy P and Lockhart, Daniel J and Milanov, Zdravko V and Morrison, Michael J and Pallares, Gabriel and Patel, Hitesh K and Pritchard, Stephanie and Wodicka, Lisa M and Zarrinkar, Patrick P. A quantitative analysis of kinase inhibitor selectivity. *Nature Biotech*, 26(1):127–132, 2008.

[46] Ismail Kola and John Landis. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*, 3(8):711 – 716, 2004.

[47] Naoki Komatsu, Kazuhiro Aoki, Masashi Yamada, Hiroko Yukinaga, Yoshihisa Fujita, Yuji Kamioka, and Michiyuki Matsuda. Development of an optimized backbone of FRET biosensors for kinases and GTPases. *Molecular Biology of the Cell*, 22:4647–4656, 2011.

[48] E Korobkova, T Emonet, J.M.G. Villar, T Shimziu, and P Cluzel. From molecular noise to behavioural variability in a single bacterium. *Nature*, 428(1), 2004.

[49] Peter O Krutzik, Janelle M Crane, Matthew R Clutter, and Garry P Nolan. High-content single-cell drug screening with phosphospecific flow cytometry. *Nat Chem Biol*, 4(2):132–142, 2008.

[50] Edo Kussell and Stanislas Leibler. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science*, 309(5743):2075–2078, 2005.

[51] I. Lestas, J. Paulsson, N.E. Ross, and G. Vinnicombe. Noise in gene regulatory

networks. *Automatic Control, IEEE Transactions on*, 53(Special Issue):189–200, Jan 2008.

[52] Hans Lineweaver and Dean Burk. The determination of enzyme dissociation constants. *Journal of the American Chemical Society*, 56(3):658–666, 1934.

[53] Thomas J. Lynch, Daphne W. Bell, Raffaella Sordella, Sarada Gurubhagavatula, Ross A. Okimoto, Brian W. Brannigan, Patricia L. Harris, Sara M. Haserlat, Jeffrey G. Supko, Frank G. Haluska, David N. Louis, David C. Christiani, Jeff Settleman, and Daniel A. Haber. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of NonSmall-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine*, 350(21):2129–2139, 2004.

[54] Harley H McAdams and Adam Arkin. Stochastic Mechanisms in Gene Expression. *PNAS*, 94:814–819, 1997.

[55] Michael Hedvat and Dennis Huszar and Andreas Herrmann and Joseph M. Gozgit and Anne Schroeder and Adam Sheehy and Ralf Buettner and David Proia and Claudia M. Kowolik and Hong Xin and Brian Armstrong and Geraldine Bebernitz and Shaobu Weng and Lin Wang and Minwei Ye and Kristen McEachern and Huawei Chen and Deborah Morosini and Kirsten Bell and Marat Alimzhanov and Stephanos Ioannidis and Patricia McCoon and Zhu A. Cao and Hua Yu and Richard Jove and Michael Zinda. The JAK2 Inhibitor AZD1480 Potently Blocks Stat3 Signaling and Oncogenesis in Solid Tumors. *Cancer Cell*, 16(6):487 – 497, 2009.

[56] Susanne Müller, Apirat Chaikuad, Nathanael S Gray, and Stefan Knapp. The ins and outs of selective kinase inhibitor development. *Nature Chem Biol*, 11(11):818–821, 2015.

[57] Leon O. Murphy, Jeffrey P. MacKeigan, and John Blenis. A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration. *Molecular and Cellular Biology*, 24(1):144–153, 2004.

[58] Mario Niepel, Sabrina L Spencer, and Peter K Sorger. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Current Opinion in Chemical Biology*, 13(56):556 – 561, 2009. Omics/Biopolymers/Model Systems.

[59] Garry P Nolan. What's wrong with drug screening today. *Nat Chem Biol*, 3(4):187,191, 2007.

[60] John J O'Shea, Apostolos Kontzias, Kunihiro Yamaoka, Yoshiya Tanaka, and Arian Laurence. Janus kinase Inhibitors in autoimmune diseases. *Ann Rheum Dis.*, 72(02):111–115, 2013.

[61] J. Guillermo Paez, Pasi A. Jnne, Jeffrey C. Lee, Sean Tracy, Heidi Greulich, Stacey Gabriel, Paula Herman, Frederic J. Kaye, Neal Lindeman, Titus J. Boggon, Katsuhiko Naoki, Hidefumi Sasaki, Yoshitaka Fujii, Michael J. Eck, William R. Sellers, Bruce E. Johnson, and Matthew Meyerson. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science*, 304(5676):1497–1500, 2004.

[62] Drew M. Pardoll. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*, 12, 2012.

[63] Johan Paulsson and Måns Ehrenberg. Random Signal Fluctuations Can Reduce Random Fluctuations in Regulated Components of Chemical Regulatory Networks. *Phys. Rev. Lett.*, 84:5447–5450, Jun 2000.

[64] Juan M. Pedraza and Alexander van Oudenaarden. Noise propagation in gene networks. *Science*, 307(5717):1965–1969, 2005.

[65] Zachary E. Perlman, Michael D. Slack, Yan Feng, Timothy J. Mitchison, Lani F. Wu, and Steven J. Altschuler. Multidimensional Drug Profiling By Automated Microscopy. *Science*, 306(5699):1194–1198, 2004.

[66] Robert J. Prill, Robert Vogel, Guillermo A. Cecchi, Grégoire Altan-Bonnet, and Gustavo Stolovitzky. Noise-driven causal inference in biomolecular networks. *PLoS ONE*, 10(6):e0125777, 06 2015.

[67] Liang Qiao, Robert B Nachbar, Ioannis G Kevrekidis, and Stanislav Y Shvartsman. Bistability and oscillations in the huang-ferrell model of mapk signaling. *PLoS Comput Biol*, 3(9):e184, 09 2007.

[68] Jonathan M Raser and Erin K O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, 2004.

[69] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science*, 308(5721):523–529, 2005.

[70] Hanna Salman, Naama Brenner, Chih-kuan Tung, Noa Elyahu, Elad Stolovicki, Lindsay Moore, Albert Libchaber, and Erez Braun. Universal protein fluctuations in populations of microorganisms. *Phys. Rev. Lett.*, 108:238105, Jun 2012.

[71] William J. Sandborn, Subrata Ghosh, Julian Panes, Ivana Vranic, Chinyu Su, Samantha Rousell, and Wojciech Niezychowski. Tofacitinib, an Oral Janus Kinase Inhibitor, in Active Ulcerative Colitis. *New England Journal of Medicine*, 367(7):616–624, 2012. PMID: 22894574.

[72] Silvia D M Santos, Peter J Verveer, and Philippe I H Bastiaens. Growth Factor-Induced MAPK Network Topology Shapes ERK Response Determining PC-12 Cell Fate. *Nature Cell Biology*, 9(3):324–30, 2007.

[73] Vahid Shahrezaei and Peter S Swain. Analytical distributions for stochastic gene expression. *PNAS*, 105(45):17256–17261, 2008.

[74] Harish Shankaran, Danielle L Ippolito, William B Chrisler, Haluk Resat, Nikki Bollinger, Lee K Opresko, and H Steven Wiley. Rapid and sustained nuclearcytoplasmic ERK oscillations induced by epidermal growth factor. *Molecular Systems Biology*, 5(1):n/a–n/a, 2009.

[75] Michael D. Slack, Elisabeth D. Martinez, Lani F. Wu, and Steven J. Altschuler. Characterizing heterogeneous cellular responses to perturbations. *Proceedings of the National Academy of Sciences*, 105(49):19306–19311, 2008.

[76] S Spencer, S Gaudet, J.G. Albeck, J.M. Burke, and P.K. Sorger. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, 459(21), 2009.

[77] John L Spudich and D.E. Koshland. Non-Genetic Individuality: Chance in the Single Cell. *Nature*, 262:467–471, 1976.

[78] Steven H. Strogatz. *Nonlinear Dynamics and Chaos with applications to Physics, Biology, Chemistry, and Engineering.* Westview Press, USA, MA, Cambridge, 2000.

[79] Gürol M Süel, Jordi Garcia-Ojalvo, Louisa M Liberman, and Michael B Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440:545–550, 2006.

[80] Gürol M. Süel, Rajan P. Kulkarni, Jonathan Dworkin, Jordi Garcia-Ojalvo, and Michael B. Elowitz. Tunability and Noise Dependence in Differentiation Dynamics. *Science*, 315(5819):1716–1719, 2007.

[81] Peter S. Swain, Michael B. Elowitz, and Eric D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[82] Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

[83] Mukund Thattai and Alexander van Oudenaarden. Attenuation of Noise in Ultrasensitive Signaling Cascades. *Biophysical Journal*, 82:2943–2950, 2002.

[84] N G Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, North-Holland, Amsterdam, 1992.

[85] Ronald F. van Vollenhoven, Roy Fleischmann, Stanley Cohen, Eun Bong Lee, Juan A. Garca Meijide, Sylke Wagner, Sarka Forejtova, Samuel H. Zwillich, David Gruben, Tamas Koncz, Gene V. Wallenstein, Sriram Krishnaswami, John D. Bradley, and Bethanie Wilkinson. Tofacitinib or Adalimumab versus Placebo in Rheumatoid Arthritis. *New England Journal of Medicine*, 367(6):508–519, 2012. PMID: 22873531.

[86] Torkel Vang, Wallace H Liu, Laurence Delacroix, Shuangding Wu, Stefan Vasile, Russell Dahl, Li Yang, Lucia Musumeci, Dana Francis, Johannes Landskron, Kjetil Tasken, Michel L Tremblay, Benedicte A Lie, Rebecca Page, Tomas Mustelin, Souad Rahmouni, Robert C Rickert, and Lutz Tautz.

LYP inhibits T-cell activation when dissociated from CSK. *Nat Chem Biol*, 8(5):437–446, 2012.

[87] Guillaume Voisinne, BrianaG. Nixon, Anna Melbinger, Georg Gasteiger, Massimo Vergassola, and Grgoire Altan-Bonnet. T Cells Integrate Local and Global Cues to Discriminate between Structurally Similar Antigens. *Cell Reports*, 11(8):1208 – 1219, 2015.

[88] Michael J. Waring, John Arrowsmith, Andrew R. Leach, Paul D. Leeson, Sam Mandrell, Robert M. Owen, Garry Pairaudeau, William D. Pennie, Stephen D. Pickett, Jibo Wang, Owen Wallace, and Alex Weir. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov*, 14(7):475,486, 2015.

[89] Hua Yu, Heehyoung Lee, Andreas Herrmann, Ralf Buettner, and Richard Jove. Revisiting STAT3 signalling in cancer: new and unexpected biological functions. *Nat Rev Cancer*, 14(11), 2014.